Journal of Statistical Physics (2024) 191:29
https://doi.org/10.1007/s10955-024-03242-7

Optimization Algorithms for Multi-species Spherical Spin Glasses

Brice Huang[1] · **Mark Sellke**[2]

Received: 14 September 2023 / Accepted: 28 January 2024 / Published online: 24 February 2024
© The Author(s) 2024

Abstract
This paper develops approximate message passing algorithms to optimize multi-species spherical spin glasses. We first show how to efficiently achieve the algorithmic threshold energy identified in our companion work (Huang and Sellke in arXiv preprint, 2023. arXiv:2303.12172), thus confirming that the Lipschitz hardness result proved therein is tight. Next we give two generalized algorithms which produce multiple outputs and show all of them are approximate critical points. Namely, in an r-species model we construct 2^r approximate critical points when the external field is stronger than a "topological trivialization" phase boundary, and exponentially many such points in the complementary regime. We also compute the local behavior of the Hamiltonian around each. These extensions are relevant for another companion work (Huang and Sellke in arXiv preprint, 2023. arXiv:2308.09677) on topological trivialization of the landscape.

Keywords Spin glasses · Optimization · Approximate message passing

1 Introduction

This paper studies the efficient optimization of a family of random non-convex functions H_N defined on high-dimensional spaces, namely the Hamiltonians of multi-species spherical spin glasses. Mean-field spin glasses have been studied since [25] as models for disordered magnetic systems and are also closely linked to random combinatorial optimization problems [12, 19, 22]. In short, their Hamiltonians are certain polynomials in many variables with independent centered Gaussian coefficients.

Communicated by Francesco Zamponi.

✉ Brice Huang
bmhuang@mit.edu

Mark Sellke
msellke@fas.harvard.edu

[1] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, USA

[2] Department of Statistics, Harvard University, Cambridge, USA

The purpose of this work is to develop efficient algorithms to optimize H_N. Our companion work [16] derives an *algorithmic threshold* ALG and proves no optimization algorithm with suitably Lipschitz dependence on H_N can achieve energy better than ALG with more than exponentially small probability. The value ALG is expressed as the maximum of a variational principle over several increasing functions, which was shown to be achieved by joining the solutions to a pair of well-posed differential equations. The first main contribution of this paper is to show that given a solution to this variational problem, so-called approximate message passing (AMP) algorithms efficiently achieve the value ALG. We note that several previous works [4, 21, 24, 26] have given similar algorithms for mean-field spin glasses with 1 species, and our algorithm is in line with the latter three.

Furthermore, we use these AMP algorithms to aid a detailed study of the landscape of H_N by probing neighborhoods of special critical points. This is related to a second companion work [17] which identifies the phase boundary for *topological trivialization* of H_N, where the number of critical points is a constant independent of N. Therein, Kac-Rice estimates are used to show that for r-species models (defined on a product of r spheres) in the "super-solvable" regime with strong external field, H_N has exactly 2^r critical points with high probability. In this paper, we give a signed AMP algorithm which explicitly approximates each of these critical points. Moreover in the complementary "sub-solvable" regime, we use AMP to construct $\exp(cN)$ separated approximate critical points with high probability. This implies the failure of *strong topological trivialization* as defined in [17], which is proved therein to hold for super-solvable models. Finally, the machinery of AMP allows us to compute the local behavior of H_N around these algorithmic outputs, giving even more precise information about the landscape.

1.1 Problem Description

Fix a finite set $\mathcal{S} = \{1, \dots, r\}$. For each positive integer N, fix a deterministic partition $\{1, \dots, N\} = \sqcup_{s\in\mathcal{S}} \mathcal{I}_s$ with $\lim_{N\to\infty} |\mathcal{I}_s|/N = \lambda_s$ where $\vec{\lambda} = (\lambda_1, \dots, \lambda_r) \in \mathbb{R}^{\mathcal{S}}_{>0}$. For $s \in \mathcal{S}$ and $\boldsymbol{x} \in \mathbb{R}^N$, let $\boldsymbol{x}_s \in \mathbb{R}^{\mathcal{I}_s}$ denote the restriction of $\boldsymbol{x}$ to coordinates $\mathcal{I}_s$. We consider the state space

$$\mathcal{B}_N = \left\{ \boldsymbol{x} \in \mathbb{R}^N : \|\boldsymbol{x}_s\|_2^2 \le \lambda_s N \quad \forall\, s \in \mathcal{S} \right\}. \tag{1.1}$$

Fix $\vec{h} = (h_1, \dots, h_r) \in \mathbb{R}^{\mathcal{S}}_{\ge 0}$ and let $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^N$. For each $k \ge 2$ fix a symmetric tensor $\Gamma^{(k)} = (\gamma_{s_1,\dots,s_k})_{s_1,\dots,s_k\in\mathcal{S}} \in (\mathbb{R}^{\mathcal{S}}_{\ge 0})^{\otimes k}$ with $\sum_{k\ge 2} 2^k \left\|\Gamma^{(k)}\right\|_\infty < \infty$, and let $\mathbf{G}^{(k)} \in (\mathbb{R}^N)^{\otimes k}$ be a tensor with i.i.d. standard Gaussian entries.

For $A \in (\mathbb{R}^{\mathcal{S}})^{\otimes k}$, $B \in (\mathbb{R}^N)^{\otimes k}$, define $A \diamond B \in (\mathbb{R}^N)^{\otimes k}$ to be the tensor with entries

$$(A \diamond B)_{i_1,\dots,i_k} = A_{s(i_1),\dots,s(i_k)} B_{i_1,\dots,i_k}, \tag{1.2}$$

where $s(i)$ denotes the $s \in \mathcal{S}$ such that $i \in \mathcal{I}_s$. Let $\boldsymbol{h} = \vec{h} \diamond \mathbf{1}$. We consider the mean-field multi-species spin glass Hamiltonian

$$H_N(\boldsymbol{\sigma}) = \langle \boldsymbol{h}, \boldsymbol{\sigma} \rangle + \widetilde{H}_N(\boldsymbol{\sigma}), \quad \text{where} \tag{1.3}$$

$$\widetilde{H}_N(\boldsymbol{\sigma}) = \sum_{k\ge 2} \frac{1}{N^{(k-1)/2}} \langle \Gamma^{(k)} \diamond \boldsymbol{G}^{(k)}, \boldsymbol{\sigma}^{\otimes k} \rangle$$

$$= \sum_{k\ge 2} \frac{1}{N^{(k-1)/2}} \sum_{i_1,\dots,i_k=1}^{N} \gamma_{s(i_1),\dots,s(i_k)} \boldsymbol{G}^{(k)}_{i_1,\dots,i_k} \sigma_{i_1} \dots \sigma_{i_k} \tag{1.4}$$

with inputs $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_N) \in \mathcal{B}_N$. For example, the choice of parameters $\Gamma^{(2)} = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ and $\Gamma^{(k)} = 0$ for $k \geq 3$ is the well-known bipartite spherical SK model [2]. For $\boldsymbol{\sigma}, \boldsymbol{\rho} \in \mathcal{B}_N$, define the species s overlap and overlap vector

$$R_s(\boldsymbol{\sigma}, \boldsymbol{\rho}) = \frac{\langle \boldsymbol{\sigma}_s, \boldsymbol{\rho}_s \rangle}{\lambda_s N}, \qquad \vec{R}(\boldsymbol{\sigma}, \boldsymbol{\rho}) = (R_1(\boldsymbol{\sigma}, \boldsymbol{\rho}), \ldots, R_r(\boldsymbol{\sigma}, \boldsymbol{\rho})) . \tag{1.5}$$

Let $\odot$ denote coordinate-wise product. For $\vec{x} = (x_1, \ldots, x_r) \in \mathbb{R}^{\mathcal{S}}$, let

$$\begin{aligned} \xi(\vec{x}) &= \sum_{k \geq 2} \langle \Gamma^{(k)} \odot \Gamma^{(k)}, (\vec{\lambda} \odot \vec{x})^{\otimes k} \rangle \\ &= \sum_{k \geq 2} \sum_{s_1, \ldots, s_k \in \mathcal{S}} \gamma^2_{s_1, \ldots, s_k} (\lambda_{s_1} x_{s_1}) \ldots (\lambda_{s_k} x_{s_k}). \end{aligned}$$

The random function $\widetilde{H}_N$ can also be described as the Gaussian process on $\mathcal{B}_N$ with covariance

$$\mathbb{E} \widetilde{H}(\boldsymbol{\sigma}) \widetilde{H}(\boldsymbol{\rho}) = N \xi(\vec{R}(\boldsymbol{\sigma}, \boldsymbol{\rho})).$$

We will also often refer to the product of spheres

$$\mathcal{S}_N = \left\{ \boldsymbol{u} \in \mathbb{R}^N \ : \ \|\boldsymbol{u}_s\|^2 = \lambda_s N \ \ \forall\, s \in \mathcal{S} \right\}. \tag{1.6}$$

It will be useful to define, for $s \in \mathcal{S}$,

$$\xi^s(\vec{x}) = \lambda_s^{-1} \partial_{x_s} \xi(\vec{x}).$$

1.2 The Value ALG

Given $(\vec{\lambda}, \xi)$, the ground state energy of the associated multi-species spherical spin glass is[1]

$$\mathsf{OPT} = \mathsf{OPT}(\xi) = \operatorname*{p\text{-}lim}_{N \to \infty} \sup_{\boldsymbol{\sigma} \in \mathcal{B}_N} H_N(\boldsymbol{\sigma})/N.$$

In the bipartite SK model mentioned above, OPT is the limiting operator norm of an IID Gaussian rectangular matrix with aspect ratio λ_1/λ_2. For large k, the asymptotic operator norm of an IID random k-tensor is similarly encoded as $\mathsf{OPT}(\xi)$ for some ξ (with e.g. $r = k$). Perhaps surprisingly, it is generally believed that polynomial-time algorithms are not in general capable of finding $\boldsymbol{\sigma} \in \mathcal{B}_N$ such that $H_N(\boldsymbol{\sigma}) \geq \mathsf{OPT}(\xi) - \varepsilon$ with high probability as $N \to \infty$. Our work [15] showed that in the single species case (and with all terms of even degree), one can identify an exact threshold ALG for the performance of a class of *Lipschitz* algorithms which includes gradient-based methods and Langevin dynamics. More recently in [16], we extended the algorithmic hardness direction of this result to multi-species spherical spin glasses, using a new proof technique that applies even when OPT is not known. The purpose of this paper is to give explicit algorithms attaining the value ALG, and we present here the formula for this value.

The algorithmic threshold ALG is given by the following variational principle. This is a simplification of the more general variational formula [16, Equation (1.7)], obtained by a partial characterization of its maximizers [16, Theorem 3]. The following generic assumption is needed therein to ensure well-posedness of the ODE (2.3) used in this description, and we will freely assume it throughout the paper.

[1] Technically the $N \to \infty$ limit is not known to exist for general ξ. Since OPT appears in the present paper only in this informal discussion, we will not belabor this point.

Assumption 1 All quadratic and cubic interactions participate in H, i.e. $\Gamma^{(2)}, \Gamma^{(3)} > 0$ coordinate-wise. We will call such models **non-degenerate**. Since this condition depends only on ξ, we similarly call ξ non-degenerate.

To optimize H_N for degenerate ξ, it suffices to apply our algorithms to a slight perturbation $\widetilde{\xi}$ which is non-degenerate and satisfies $\|\xi - \widetilde{\xi}\|_{C^3([0,1]^r)} \le \varepsilon$ to obtain the guarantees in this and the next section. Here, $C^3([0,1]^r)$ denotes the norm

$$\|\xi\|_{C^3([0,1]^r)} = \sup_{\vec{x}\in[0,1]^r} \max\left\{|\xi(\vec{x})|, \|\nabla\xi(\vec{x})\|_\infty, \|\nabla^2\xi(\vec{x})\|_\infty, \|\nabla^3\xi(\vec{x})\|_\infty\right\}.$$

Since both the ground state and the more general ALG formula in [16] (allowing degenerate ξ) vary continuously in ξ, there is essentially no loss of generality in assuming non-degeneracy.

The formula for ALG is described by two cases depending on whether $\vec{1} = 1^{\mathcal{S}}$ is super-solvable as defined below.

Definition 1.1 A matrix $M \in \mathbb{R}^{\mathcal{S}\times\mathcal{S}}$ is **diagonally signed** if $M_{i,i} \ge 0$ and $M_{i,j} < 0$ for all $i \ne j$.

Definition 1.2 A symmetric diagonally signed matrix M is **super-solvable** if it is positive semidefinite, and **solvable** if it is furthermore singular; otherwise M is **strictly sub-solvable**. A point $\vec{x} \in (0,1]^{\mathcal{S}}$ is super-solvable, solvable, or strictly sub-solvable if $M^*(\vec{x})$ is, where

$$M^*(\vec{x}) = \operatorname{diag}\left(\left(\frac{\partial_{x_s}\xi(\vec{x}) + \lambda_s h_s^2}{x_s}\right)_{s\in\mathcal{S}}\right) - \left(\partial_{x_s,x_{s'}}\xi(\vec{x})\right)_{s,s'\in\mathcal{S}}. \tag{1.7}$$

We also adopt the convention that $\vec{0}$ is always super-solvable, and solvable if $\vec{h} = \vec{0}$.

The following will be useful.

Proposition 1.3 ([16, Proposition 4.3], see also [17, Lemma 2.5]) *If the square matrix M is diagonally signed, then the minimal eigenvalue $\lambda_{\min}(M)$ has multiplicity* 1*, and the corresponding eigenvector $\vec{v}$ has strictly positive entries. Moreover*

$$\lambda_{\min}(M) = \sup_{\vec{v}\succ\vec{0}} \min_{s\in\mathcal{S}} \frac{(M\vec{v})_s}{v_s},$$

and the supremum is uniquely attained at $\vec{v}$.

It is easy to see that any $x \in (0,1]^{\mathcal{S}}$ is sub-solvable when $\vec{h} = \vec{0}$, and that super-solvability is a coordinate-wise increasing property of $\vec{h}$. For our purposes, an external field is large if $\vec{1}$ is super-solvable and small if $\vec{1}$ is strictly sub-solvable. (Unfortunately we do not have more refined intuition for the precise form of M^* above, nor the resulting phase boundary between super and sub-solvability.) As shown in our companion work [17], in super-solvable models the external fields $\boldsymbol{h}$ are strong enough to trivialize the "glassy" nature of the landscape for H_N. Namely the number of critical points is exactly 2^r with high probability, the minimum number of any generic smooth ("Morse") function on a product of r spheres. By contrast in the sub-solvable case, the expected number of critical points is exponentially large in the dimension N. As explained below, the optimization algorithms are also simpler in the super-solvable case.

Definition 1.4 (*Algorithmic Threshold, Super-Solvable Case*) If $\vec{1}$ is super-solvable, then

$$\mathsf{ALG} = \sum_{s\in\mathcal{S}} \lambda_s \sqrt{\xi^s(\vec{1}) + h_s^2}.$$

When $\vec{1}$ is strictly sub-solvable, the formula for ALG becomes more complicated and depends on the optimal choice of a increasing C^2 function $\Phi : [q_1, 1] \to [0, 1]^{\mathcal{S}}$ satisfying certain conditions. We term such Φ *pseudo-maximizers* and defer the formal definition to Definition 2.1. Note that $q_1 \in [0, 1]$ is not fixed, but is determined by the choice of Φ.

Definition 1.5 (*Algorithmic Threshold, Strictly Sub-solvable Case*) If $\vec{1}$ is strictly sub-solvable, then with the maximum taken over all pseudo-maximizers Φ of $\mathbb{A}$,

$$\begin{aligned} \mathsf{ALG} &= \max_{\Phi} \mathbb{A}(\Phi); \\ \mathbb{A}(\Phi) &\equiv \sum_{s \in \mathcal{S}} \lambda_s \left[\sqrt{\Phi_s(q_1)(\xi^s(\Phi(q_1)) + h_s^2)} + \int_{q_1}^1 \sqrt{\Phi_s'(q)(\xi^s \circ \Phi)'(q)} \, \mathrm{d}q \right]. \end{aligned} \tag{1.8}$$

See [16, Remark 1.3] for an approach to maximizing $\mathbb{A}$ using the well-posedness of the ODEs (2.2), (2.3) in the definition of pseudo-maximizer. The computational complexity of this task is in particular independent of N.

The following theorem is our main result. We equip the space $\mathcal{H}_N$ of Hamiltonians H_N with the following distance. We identify H_N with its disorder coefficients $(\boldsymbol{G}^{(k)})_{k \geq 2}$, which we arrange in an arbitrary but fixed order into an infinite vector $\boldsymbol{g}(H_N)$, and define

$$\| H_N - H_N' \|_2 = \| \boldsymbol{g}(H_N) - \boldsymbol{g}(H_N') \|_2.$$

(In other words, $\| H_N - H_N' \|_2^2$ is the sum of squared differences $(g_{i_1,\ldots,i_k} - g'_{i_1,\ldots,i_k})^2$ between all corresponding pairs of coefficients in $(\boldsymbol{G}^{(k)})_{k \geq 2}$ and $(\boldsymbol{G}'^{(k)})_{k \geq 2}$.) We say an algorithm $\mathcal{A}_N : \mathcal{H}_N \to \mathcal{B}_N$ is τ-Lipschitz if

$$\| \mathcal{A}_N(H_N) - \mathcal{A}_N(H_N') \|_2 \leq \tau \| H_N - H_N' \|_2, \qquad \forall H_N, H_N' \in \mathcal{H}_N.$$

Note that $\| H_N - H_N' \|_2$ may be infinite, and if so this condition holds vacuously for such pairs (H_N, H_N'). Here and throughout, all implicit constants may depend also on $(\xi, \vec{h}, \vec{\lambda})$.

Theorem 1 *For any $\varepsilon > 0$, there exists an $O_\varepsilon(1)$-Lipschitz $\mathcal{A}_N : \mathcal{H}_N \to \mathcal{B}_N$ such that*

$$\mathbb{P}[H_N(\mathcal{A}_N(H_N))/N \geq \mathsf{ALG} - \varepsilon] \geq 1 - \exp(-cN), \quad c = c(\varepsilon) > 0.$$

The main result in our companion work [16, Theorem 1] states that any τ-Lipschitz $\mathcal{A}_N : \mathcal{H}_N \to \mathcal{B}_N$ satisfies, for the same threshold ALG and N sufficiently large,

$$\mathbb{P}[H_N(\mathcal{A}_N(H_N))/N \geq \mathsf{ALG} + \varepsilon] \leq \exp(-cN), \quad c = c(\varepsilon, \tau) > 0.$$

Together these results thus characterize the best possible Lipschitz optimization algorithms for multi-species spherical spin glasses.

We prove Theorem 1 with an explicit algorithm based on AMP, following a recent line of work [4, 5, 21, 24, 26]. Such algorithms are shown to be Lipschitz (up to modification on a set with $\exp(-cN)$ probability) in [15, Sect. 8]. AMP algorithms also have computational complexity which is linear in the input size when H_N is a polynomial of finite degree (modulo solving for Φ, a task that does not depend on N). See [4, Remark 2.1] for related discussion on this last point.

Similarly to [5, 24], our algorithm has two phases, a "root-finding" phase and a "tree-descending" phase. Roughly speaking, the set of points reachable by our algorithm has the geometry of a densely branching ultrametric tree, which is rooted at the origin when $\boldsymbol{h} = \boldsymbol{0}$ and more generally at a random point correlated with $\boldsymbol{h}$. The first phase identifies this root, and the second traces a root-to-leaf path of this tree. The structure of the first phase is similar

to the original AMP algorithm of [9] for the SK model at high-temperature, while the latter *incremental* AMP technique was introduced in [21].

For the purposes of this paper, the significance of (super, sub)-solvability is as follows. When the external field is sufficiently large, the root moves all the way to the boundary of $\mathcal{B}_N$ (in all r species) and the algorithmic tree becomes degenerate. In [16], it is shown that the external field is large enough for this to occur if and only if $\vec{1}$ is super-solvable. Moreover, [17] shows this condition coincides with strong topological trivialization (defined therein) of the optimization landscape.

In Sect. 3 we extend our main algorithm in several ways. In Sect. 3.1 we define 2^r signed generalizations of the root-finding algorithm with similar behavior. In Sect. 3.2 we compute the gradients of H_N at the points output by our algorithm, in both cases when $\vec{1}$ is super-solvable and sub-solvable. In particular, we show that they are approximate critical points on the product of spheres $\mathcal{S}_N$ (defined in (1.6)). As explained in Remark 3.1, in the strictly super-solvable case these 2^r outputs approximate the 2^r genuine critical points of H_N on $\mathcal{S}_N$. The sub-solvable case of this computation is used in our companion paper [17, Theorem 1.5(c) and Sect. 5.3] to show failure of *annealed* topological trivialization in the sub-solvable case. Finally in Sect. 3.3 we give a modification of the tree-descending phase for the super-solvable case. It constructs $\exp(cN)$ well-separated approximate critical points arranged in a densely branching ultrametric tree; this implies the failure of *strong* topological trivialization in [17, Definition 6 and Theorem 1.6].

1.3 Notations

Throughout, we will use boldface lowercase letters $(\boldsymbol{u}, \boldsymbol{v}, \ldots)$ to denote vectors in $\mathbb{R}^N$, and lowercase letters with vector sign $(\vec{u}, \vec{v}, \ldots)$ to denote vectors in $\mathbb{R}^{\mathcal{S}} \simeq \mathbb{R}^r$. Similarly, boldface uppercase letters denote matrices or tensors in $(\mathbb{R}^N)^{\otimes k}$, and non-boldface uppercase letters denote matrices or tensors in $(\mathbb{R}^r)^{\otimes k}$. We let

$$\langle \boldsymbol{v} \rangle_N = N^{-1} \sum_{i \le N} v_i; \qquad \langle \boldsymbol{u}, \boldsymbol{v} \rangle_N = N^{-1} \sum_{i \le N} u_i v_i = \langle \vec{\lambda}, \vec{R}(\boldsymbol{u}, \boldsymbol{v}) \rangle$$

for $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^N$. The corresponding norm is

$$\|\boldsymbol{u}\|_N = \langle \boldsymbol{u}, \boldsymbol{u} \rangle_N^{1/2} = \sqrt{\sum_s \lambda_s R_s(\boldsymbol{u}, \boldsymbol{u})}.$$

Next $a_N \simeq b_N$ means that $a_N - b_N$ converges in probability to 0. Analogously, for two vectors $\boldsymbol{u}_N, \boldsymbol{v}_N$, we write $\boldsymbol{u}_N \simeq \boldsymbol{v}_N$ when $\|\boldsymbol{u}_N - \boldsymbol{v}_N\|_N$ converges in probability to 0. We denote limits in probability by p-$\lim_{N \to \infty}$. Analogously we write $\approx_\delta$ to denote asymptotic equality as $\delta \to 0$.

For any tensor $\boldsymbol{A} \in (\mathbb{R}^N)^{\otimes k}$, we define the operator norm

$$\|\boldsymbol{A}\|_{\mathrm{op}} = \sup_{\|\boldsymbol{\sigma}^1\|, \ldots, \|\boldsymbol{\sigma}^k\| \le 1} \left| \langle \boldsymbol{A}, \boldsymbol{\sigma}^1 \otimes \cdots \otimes \boldsymbol{\sigma}^k \rangle \right|.$$

The following proposition shows that with exponentially good probability, the operator norms of all constant-order gradients of H_N are bounded on the appropriate scale.

Proposition 1.6 ([16, Proposition 1.13]) *For any fixed model $(\xi, \vec{h})$ there exists a constant $c > 0$, sequence $(K_N)_{N \ge 1}$ of convex sets $K_N \subseteq \mathcal{H}_N$, and sequence of constants $(C_k)_{k \ge 1}$ independent of N, such that the following properties hold.*

(a) $\mathbb{P}[H_N \in K_N] \geq 1 - e^{-cN}$;
(b) *For all* $H_N \in K_N$ *and* $\boldsymbol{x} \in \mathcal{B}_N$,

$$\left\| \nabla^k H_N(\boldsymbol{x}) \right\|_{op} \leq C_k N^{1-\frac{k}{2}}. \tag{1.9}$$

2 Achieving Energy ALG

In this section we prove Theorem 1 by exhibiting an AMP algorithm. Throughout this section, Assumption 1 on non-degeneracy of ξ will be enforced without loss of generality.

2.1 Definition of Pseudo-Maximizer

As mentioned before Definition 1.5, the threshold ALG in the sub-solvable case depends on a notion of pseudo-maximizer. We now provide this definition, which was derived in [16, Theorem 3] as a necessary condition for Φ to maximize $\mathbb{A}$ defined in (1.8) (and it is proved therein that a maximizer always exists).

Definition 2.1 A coordinate-wise strictly increasing C^2 function $\Phi : [q_1, 1] \to [0, 1]^{\mathcal{S}}$, for some $q_1 \in [0, 1]$, is a pseudo-maximizer if:

(1) Φ is *admissible*, meaning it satisfies the normalization

$$\langle \vec{\lambda}, \Phi(q) \rangle = q, \quad \forall q \in [q_1, 1]. \tag{2.1}$$

In particular $\Phi(1) = \vec{1}$.
(2) $\Phi(q_1)$ is solvable.
(3) The derivative at q_1 satisfies $M^*(\Phi(q_1))\Phi'(q_1) = \vec{0}$. This amounts to no restriction when $\vec{h} = \vec{0}$ and thus $(q_1, \Phi(q_1)) = (0, \vec{0})$; when $\vec{h} \neq \vec{0}$ it means that

$$\Phi_s'(q_1) = \frac{\Phi_s(q_1)(\xi^s \circ \Phi)'(q_1)}{\xi^s(\Phi(q_1)) + h_s^2}, \quad s \in \mathcal{S}. \tag{2.2}$$

(4) For all $q \in [q_1, 1]$, Φ solves the (second-order) *tree-descending* differential equation:

$$\Psi(q) \equiv \frac{1}{\Phi_s'(q)} \frac{\mathrm{d}}{\mathrm{d}q} \sqrt{\frac{\Phi_s'(q)}{(\xi^s \circ \Phi)'(q)}} \tag{2.3}$$

is independent of the species s. (See [16, Lemma 4.37] for well-posedness of this ODE.)

Note that there may exist multiple such Φ, see [16, Figure 2]. If $\vec{1}$ is super-solvable, we adopt the convention that $q_1 = 1$ and Φ has domain $\{1\}$.

We now give an efficient AMP algorithm achieving energy $\mathbb{A}(\Phi)$ for any pseudo-maximizer Φ. In particular for the optimal pseudo-maximizer this achieves energy ALG.

2.2 Review of Approximate Message Passing

Here we recall the class of AMP algorithms, specialized to our setting of interest. We initialize AMP with a deterministic vector $\boldsymbol{w}^0$ with coordinates

$$w_i^0 = w_{s(i)} \tag{2.4}$$

depending only on the species. Let $f_{t,s} : \mathbb{R}^{t+1} \to \mathbb{R}$ be a Lipschitz function for each $(t, s) \in \mathbb{Z}_{\geq 0} \times \mathcal{S}$. For $(\boldsymbol{w}^0, \boldsymbol{w}^1, \ldots, \boldsymbol{w}^t) \in \mathbb{R}^{N\times(t+1)}$, let $f_t(\boldsymbol{w}^0, \boldsymbol{w}^1, \ldots, \boldsymbol{w}^t) \in \mathbb{R}^N$ be given by

$$f_t(\boldsymbol{w}^0, \boldsymbol{w}^1, \ldots, \boldsymbol{w}^t)_i = f_{t,s(i)}(w_i^1, \ldots, w_i^t), \quad i \in [N].$$

We generate subsequent iterates through recursions of the following form, where $\mathbf{ons}_t$ is known as the *Onsager correction term*:

$$\begin{aligned} \boldsymbol{w}^{t+1} &= \nabla H_N(\boldsymbol{m}^t) - \mathbf{ons}_t; \\ \boldsymbol{m}^t &= f_t(\boldsymbol{w}^0, \boldsymbol{w}^1, \ldots, \boldsymbol{w}^t); \qquad (2.5) \\ \mathbf{ons}_t &= \sum_{t' \leq t} d_{t,t'} \diamond f_{t'-1}(\boldsymbol{w}^1, \ldots, \boldsymbol{w}^{t'-1}); \qquad (2.6) \end{aligned}$$

$$d_{t,t',s} = \left(\sum_{s' \in \mathcal{S}} \partial_{x_{s'}} \xi^s \left(\left(\mathbb{E}[M_{s''}^t M_{s''}^{t'-1}] \right)_{s'' \in \mathcal{S}} \right) \cdot \mathbb{E}\left[\partial_{W_{s'}^{t'}} f_{t,s'}(W_{s'}^0, \ldots, W_{s'}^t) \right] \right). \qquad (2.7)$$

Here W_s^t, M_s^t are defined as follows. $W_s^0 = w_s$ and the variables $(\widetilde{W}_s^t)_{(t,s)\in\mathbb{Z}_{\geq 1}\times\mathcal{S}}$ form a centered Gaussian process with covariance defined recursively by

$$\begin{aligned} \mathbb{E}[\widetilde{W}_s^{t+1} \widetilde{W}_s^{t'+1}] &= \xi^s \left(\mathbb{E}[f_{t,s}(W_s^0, \ldots, W_s^t) f_{t',s}(W_s^0, \ldots, W_s^{t'})] \right), \\ W_s^t &= \widetilde{W}_s^t + h_s; \\ M_s^t &= f_{t,s}(W_s^0, \ldots, W_s^t) \end{aligned} \qquad (2.8)$$

and $\mathbb{E}[\widetilde{W}_s^{t+1} \widetilde{W}_{s'}^{t'+1}] = 0$ if $s \neq s'$ (i.e. different species are independent).

The following *state evolution* characterizes the behavior of the above iterates. It states that for each $s \in \mathcal{S}$, when $i \in \mathcal{I}_s$ is uniformly random the sequence of coordinates $(w_i^1, w_i^2, \ldots, w_i^t)$ has the same law as $(W_s^1, \ldots, W_s^t)$. Say a function $\psi : \mathbb{R}^\ell \to \mathbb{R}$ is pseudo-Lipschitz if $|\psi(x) - \psi(y)| \leq C(1 + |x| + |y|)|x - y|$ for a constant C.

Proposition 2.2 *For any pseudo-Lipschitz function ψ and $\ell \in \mathbb{Z}_{\geq 0}$, $s \in \mathcal{S}$,*

$$\operatorname*{p-lim}_{N\to\infty} \frac{1}{N_s} \sum_{i \in \mathcal{I}_s} \psi(\boldsymbol{w}_i^0, \ldots, \boldsymbol{w}_i^\ell) = \mathbb{E}[\psi(W_s^0, \ldots, W_s^\ell)]. \qquad (2.9)$$

This proposition allows us to read off normalized inner products of the AMP iterates, since e.g.

$$\langle \boldsymbol{w}^k, \boldsymbol{w}^\ell \rangle_N \simeq \sum_{s \in \mathcal{S}} \lambda_s \mathbb{E}[W_s^k W_s^\ell].$$

Proposition 2.2 is proved in Appendix 1. In fact we show a slight generalization allowing $f_t = f_t(\boldsymbol{w}^0, \ldots, \boldsymbol{w}^t, \boldsymbol{g}^0, \ldots, \boldsymbol{g}^t)$ to depend also on independently generated vectors $(\boldsymbol{g}^0, \ldots, \boldsymbol{g}^t) \in \mathbb{R}^{N(t+1)}$. When using this extension, we will always take each $\boldsymbol{g}^t \sim \mathcal{N}(0, I_N)$ to be standard Gaussian. The more general result essentially says that $\boldsymbol{g}_t$ still acts as an independent Gaussian for the purposes of state evolution. Since this is relatively intuitive, we refer to Theorem 2 in the appendix for a precise statement.

For random matrices (i.e. the case of quadratic H) there is a considerable literature establishing state evolution in many settings beginning with [7, 9] and later [6, 8, 10, 11, 13] (see also [14] for a survey of many statistical applications). The generalization to tensors was introduced in [23] and proved in [4], whose approach we follow.

2.3 Stage I: Finding the Root of the Ultrametric Tree

Our goal in this subsection will be to compute a vector $\boldsymbol{m}^{\underline{\ell}}$ satisfying

$$\operatorname*{p\text{-}lim}_{\underline{\ell}\to\infty} \lim_{N\to\infty} \vec{R}(\boldsymbol{m}^{\underline{\ell}}, \boldsymbol{m}^{\underline{\ell}}) = \Phi(q_1)$$

and with the correct energy value (as stated in Lemma 2.5). We take as given a maximizer Φ to $\mathbb{A}$ with domain $[q_1, 1]$. Recall $\Phi(q_1)$ is super-solvable: either $\vec{1}$ is strictly sub-solvable, in which case $\Phi(q_1)$ is solvable, or $\vec{1}$ is super-solvable, in which case $\Phi(q_1) = \Phi(1) = \vec{1}$.

We use the initialization

$$w_i^0 = \sqrt{\xi^s(\Phi(q_1)) + h_s^2}, \quad i \in \mathcal{I}_s.$$

Define the vector $\vec{a} \in \mathbb{R}^{\mathcal{S}}$ by

$$a_s = \sqrt{\frac{\Phi_s(q_1)}{\xi^s(\Phi(q_1)) + h_s^2}}.$$

Subsequent iterates are defined via the following recursion.

$$\begin{aligned} \boldsymbol{w}^{k+1} &= \nabla H_N(\boldsymbol{m}^k) - \vec{b}_k \diamond \boldsymbol{m}^{k-1} \\ &= \boldsymbol{h} + \nabla \widetilde{H}_N(\boldsymbol{m}^k) - \vec{b}_k \diamond \boldsymbol{m}^{k-1}; \qquad (2.10) \\ \boldsymbol{m}^k &= \vec{a} \diamond \boldsymbol{w}^k \qquad (2.11) \\ b_{k,s} &\equiv \sum_{s' \in \mathcal{S}} a_{s'} \partial_{s'} \xi^s \big(\vec{R}(\boldsymbol{m}^k, \boldsymbol{m}^{k-1})\big). \qquad (2.12) \end{aligned}$$

The last term in (2.10) comes from specializing the formula (2.6) for the Onsager term.

Next recalling (2.8), let $(W_s^j, M_s^j)_{j\geq 0, s\in\mathcal{S}}$ be the state evolution limit of the coordinates of

$$(\boldsymbol{w}^0, \boldsymbol{m}^0, \ldots, \boldsymbol{w}^k, \boldsymbol{m}^k)$$

as $N \to \infty$. Concretely, each W_s^j is Gaussian with mean h_s and

$$M_s^j = \sqrt{\frac{\Phi_s(q_1)}{\xi^s(\Phi(q_1)) + h_s^2}} \cdot W_s^j, \quad j \geq 0,\ s \in \mathcal{S}.$$

We next compute the covariance of the Gaussians $\widetilde{W}_s^j = W_s^j - h_s$. Define $\vec{\alpha} : \mathbb{R}^{\mathcal{S}}_{\geq 0} \to \mathbb{R}^{\mathcal{S}}_{\geq 0}$ by

$$\alpha_s(\vec{x}) = \big(\xi^s(\vec{x}) + h_s^2\big)\left(\frac{\Phi_s(q_1)}{\xi^s(\Phi(q_1)) + h_s^2}\right). \qquad (2.13)$$

Define the (deterministic) $\mathbb{R}^{\mathcal{S}}_{\geq 0}$-valued sequence $(\vec{R}^0, \vec{R}^1, \ldots)$ of asymptotic overlaps recursively by $\vec{R}^0 = \vec{0}$ and $\vec{R}^{k+1} = \vec{\alpha}(\vec{R}^k)$.

Lemma 2.3 *For integers $0 \leq j < k$, the following equalities hold (the first in distribution):*

$$\begin{aligned} W_s^j &\stackrel{d}{=} h_s + Z\sqrt{\xi^s(\Phi(q_1))}, \quad Z \sim \mathcal{N}(0,1) \qquad (2.14) \\ \mathbb{E}[\widetilde{W}_s^j \widetilde{W}_s^k] &= \xi^s(\vec{R}^j) \qquad (2.15) \\ \mathbb{E}[(M_s^j)^2] &= \Phi_s(q_1) \qquad (2.16) \\ \mathbb{E}[M_s^j M_s^k] &= R_s^{j+1}. \qquad (2.17) \end{aligned}$$

Proof We proceed by induction on j, first showing (2.14) and (2.16) together. As a base case, (2.14) holds for $j = 0$ by initialization. For the inductive step, assume first that (2.14) holds for j. Then by the definition (2.11),

$$\begin{aligned}\mathbb{E}\left[(M_s^j)^2\right] &= \left(\xi^s(\Phi(q_1)) + h_s^2\right) \cdot a_s^2 \\ &= \left(\xi^s(\Phi(q_1)) + h_s^2\right) \cdot \left(\frac{\Phi_s(q_1)}{\xi^s(\Phi(q_1)) + h_s^2}\right) \\ &= \Phi_s(q_1)\end{aligned}$$

so that (2.14) implies (2.16) for each $j \geq 0$. On the other hand, state evolution directly implies that if (2.16) holds for j then (2.14) holds for $j + 1$. This establishes (2.14) and (2.16) for all $j \geq 0$.

We similarly show (2.15) and (2.17) together by induction, beginning with (2.15). When $j = 0$ it is clear because $\widetilde{W}_s^k$ is mean zero and independent of $\widetilde{W}_s^0$. Just as above, it follows from state evolution that (2.15) for (j, k) implies (2.17) for (j, k) which in turn implies (2.15) for $(j + 1, k + 1)$. Hence induction on j proves (2.15) and (2.17) for all (j, k). □

The next lemma is crucial and uses super-solvability of $\Phi(q_1)$.

Lemma 2.4 *The limit* $\vec{R}^\infty \equiv \lim_{j\to\infty} \vec{R}^j$ *exists and equals* $\Phi(q_1)$.

Proof First we observe that $\vec{\alpha}$ (recall (2.13)) is coordinate-wise strictly increasing in the sense that if $0 \preceq x \prec y$ then $\vec{\alpha}(x) \prec \vec{\alpha}(y)$. Moreover $\vec{\alpha}(\vec{0}) \succ 0$ (assuming $\vec{h} \neq 0$, else the result is trivial) and $\vec{\alpha}(\Phi(q_1)) = \Phi(q_1)$. Therefore $\vec{R}^\infty$ exists, $\vec{\alpha}(\vec{R}^\infty) = \vec{R}^\infty$, and

$$\vec{0} \preceq \vec{R}^\infty \preceq \Phi(q_1).$$

It remains to show that the above forces $\vec{R}^\infty = \Phi(q_1)$ to hold.

Let $M \in \mathbb{R}^{\mathcal{S}\times\mathcal{S}}$ be the matrix with entries $M_{s,s'} = \frac{\mathrm{d}}{\mathrm{d}t}\vec{\alpha}_s(\Phi(q_1) + te_{s'})|_{t=0}$ for $e_{s'}$ a standard basis vector. Then M is the derivative matrix for $\vec{\alpha}$ at $\Phi(q_1)$ in the sense that for any $\vec{u} \in \mathbb{R}^{\mathcal{S}}$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\vec{\alpha}(\Phi(q_1) + t\vec{u})|_{t=0} = M\vec{u}.$$

We easily calculate that

$$M_{s,s'} = \frac{\Phi_s(q_1)\partial_{x_s,x_{s'}}\xi(\Phi(q_1))}{\partial_{x_s}\xi(\Phi(q_1)) + \lambda_s h_s^2}.$$

We claim that for any entry-wise non-negative vector $\vec{w} \in \mathbb{R}^{\mathcal{S}}_{\geq 0}$,

$$(M\vec{w})_s \leq w_s \tag{2.18}$$

for some $s \in \mathcal{S}$. Indeed, suppose to the contrary that $(M\vec{w})_s > w_s$ for all $s \in \mathcal{S}$. This rearranges to

$$\frac{\partial_{x_s}\xi(\Phi(q_1)) + \lambda_s h_s^2}{\Phi_s(q_1)} w_s - \sum_{s'\in\mathcal{S}} \partial_{x_s,x_{s'}}\xi(\Phi(q_1))w_{s'} < 0 \quad \forall s \in \mathcal{S},$$

i.e. $M^*(\Phi(q_1))\vec{w} \prec \vec{0}$ (recall (1.7)). Proposition 1.3 then implies that $\boldsymbol{\lambda}_{\min}(M^*(\Phi(q_1))) < 0$, so $\Phi(q_1)$ is strictly sub-solvable, which is a contradiction. Thus (2.18) holds for some $s \in \mathcal{S}$.

Now suppose for sake of contradiction that $\vec{R}^\infty \prec \Phi(q_1)$, let $\vec{w} = \Phi(q_1) - \vec{R}^\infty$, and choose $s \in \mathcal{S}$ such that (2.18) holds. Write $f(t) = \alpha_s(\Phi(q_1) + t\vec{w})$. Since α_s is a polynomial

with non-negative coefficients and ξ is non-degenerate, f is strictly convex and strictly increasing on $[-1, 0]$. Hence

$$\alpha_s(\vec{R}^\infty) = f(-1) > f(0) - f'(0) \geq \Phi_s(q_1) - (M\vec{w})_s \overset{(2.18)}{\geq} \Phi_s(q_1) - w_s = R_s^\infty.$$

The first inequality above is strict, so we deduce that $\vec{\alpha}(\vec{R}^\infty) \neq \vec{R}^\infty$ if $\vec{R}^\infty \prec \Phi(q_1)$. This contradicts the definition of $\vec{R}^\infty$. Therefore $\vec{R}^\infty = \Phi(q_1)$, completing the proof. □

Remark 2.1 Super-solvability of $\Phi(q_1)$ is a tight condition for the above argument to hold, as the matrix M above needs to have Perron-Frobenius eigenvalue at most 1. Indeed suppose that $\Phi(q_1)$ was chosen so that $\lambda_1(M) > 1$. Then there exists $\vec{w} \in \mathbb{R}^{\mathcal{S}}_{>0}$ with $M\vec{w} \succ \vec{w}$. Letting $\vec{x} = \Phi(q_1) - \varepsilon\vec{w}$ for small $\varepsilon > 0$, we find $\vec{\alpha}(\vec{x}) \prec \vec{x}$. Monotonicity implies that $\vec{\alpha}$ maps the compact, convex set

$$K = \{\vec{y} \in [0, 1]^{\mathcal{S}} \ : \ \vec{0} \preceq \vec{y} \preceq \vec{x}\}$$

into itself. By the Brouwer fixed point theorem, a fixed point of $\vec{\alpha}$ strictly smaller than $\Phi(q_1)$ exists whenever $\Phi(q_1)$ is strictly subsolvable.

We finish our analysis of the first AMP phase by computing the asymptotic energy it achieves. As expected, the resulting value agrees with the first term in the formula (1.8) for ALG.

Lemma 2.5

$$\lim_{k\to\infty} \text{p-lim}_{N\to\infty} \frac{H_N(\boldsymbol{m}^k)}{N} = \sum_{s\in\mathcal{S}} \lambda_s \sqrt{\Phi_s(q_1) \cdot \left(h_s^2 + \xi^s(\Phi(q_1))\right)}.$$

Proof We use the identity

$$\frac{H_N(\boldsymbol{m}^k)}{N} = \langle \boldsymbol{h}, \boldsymbol{m}^k \rangle_N + \int_0^1 \langle \boldsymbol{m}^k, \nabla \widetilde{H}_N(t\boldsymbol{m}^k) \rangle_N \mathrm{d}t \tag{2.19}$$

and interchange the limit in probability with the integral. To compute $\text{p-lim}_{N\to\infty} \langle \boldsymbol{m}^k, \nabla \widetilde{H}_N(t\boldsymbol{m}^k) \rangle$ we introduce an auxiliary AMP step

$$\boldsymbol{y}^{k+1} = \nabla \widetilde{H}_N(t\boldsymbol{m}^k) - t\vec{b}_k \diamond \boldsymbol{m}^{k-1}$$

which depends implicitly on $t \in [0, 1]$. Rearranging yields

$$\begin{aligned} \vec{R}(\boldsymbol{m}^k, \nabla \widetilde{H}_N(t\boldsymbol{m}^k)) &= \vec{R}(\boldsymbol{m}^k, \boldsymbol{y}^{k+1}) + t \cdot \left(\vec{R}(\boldsymbol{m}^k, \boldsymbol{m}^{k-1}) \odot \vec{b}_k\right) \\ &\simeq \vec{R}(\boldsymbol{m}^k, \boldsymbol{y}^{k+1}) + t \cdot \left(\vec{R}^k \odot \vec{b}_k\right). \end{aligned}$$

For the first term, recalling (2.11) yields

$$\begin{aligned} R_s(\boldsymbol{m}^k, \boldsymbol{y}^{k+1}) &= \mathbb{E}[a_s W_s^k Y_s^{k+1}] \\ &= a_s \ \xi^s(t\vec{R}^k). \end{aligned}$$

Note also that

$$\lambda_s \partial_{s'} \xi^s(\vec{R}^k) = \partial_{x_s, x_{s'}} \xi(\vec{R}^k) = \lambda_{s'} \partial_s \xi^{s'}(\vec{R}^k). \tag{2.20}$$

Integrating with respect to t, and switching the roles of s, s' in applying (2.20), we thus find

$$\begin{aligned}\int_0^1 \langle \boldsymbol{m}^k, \nabla \widetilde{H}_N(t\boldsymbol{m}^k)\rangle_N \mathrm{d}t &\simeq \sum_{s\in\mathcal{S}} \lambda_s \int_0^1 R_s(\boldsymbol{m}^k, \nabla \widetilde{H}_N(t\boldsymbol{m}^k))\mathrm{d}t \\ &\simeq \sum_{s\in\mathcal{S}} \lambda_s \int_0^1 \Big(a_s \xi^s(t\vec{R}^k) + t R_s^k \sum_{s'} a_{s'} \partial_{s'} \xi^s(\vec{R}^k)\Big)\,\mathrm{d}t \\ &\overset{(2.20)}{=} \sum_{s\in\mathcal{S}} \lambda_s \int_0^1 \Big(a_s \xi^s(t\vec{R}^k) + t a_s \sum_{s'} R_{s'}^k \partial_{s'} \xi^s(\vec{R}^k)\Big)\,\mathrm{d}t \\ &= \sum_{s\in\mathcal{S}} \lambda_s a_s \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\Big(t\,\xi^s(t\,\vec{R}^k)\Big)\,\mathrm{d}t \\ &= \sum_{s\in\mathcal{S}} \lambda_s a_s \xi^s(\vec{R}^k).\end{aligned}$$

Finally the external field $\boldsymbol{h}$ gives energy contribution

$$\langle \boldsymbol{h}, \boldsymbol{m}^k\rangle_N \simeq \sum_{s\in\mathcal{S}} \lambda_s h_s \mathbb{E}[M_s^k] = \sum_{s\in\mathcal{S}} \lambda_s a_s h_s^2.$$

Since $\vec{R}^\infty = \Phi(q_1)$ by Lemma 2.4, we conclude

$$\begin{aligned}\lim_{k\to\infty} \operatorname*{p-lim}_{N\to\infty} \frac{H_N(\boldsymbol{m}^k)}{N} &= \sum_{s\in\mathcal{S}} \lambda_s a_s \big(h_s^2 + \xi^s(\Phi(q_1))\big) \\ &= \sum_{s\in\mathcal{S}} \lambda_s \sqrt{\Phi_s(q_1)\cdot\big(h_s^2 + \xi^s(\Phi(q_1))\big)}.\end{aligned}$$

□

2.4 Stage II: Descending the Ultrametric Tree

We now turn to the second phase which uses incremental approximate message passing. Choose a large integer $\underline{\ell}$, and with $\delta = \underline{\ell}^{-1}$ let

$$q_\ell^\delta = q_1 + (\ell - \underline{\ell})\delta, \quad \ell \geq 0.$$

We then define

$$\boldsymbol{n}^{\underline{\ell}} = \boldsymbol{m}^{\underline{\ell}} + \sqrt{\Phi(q_1+\delta) - \Phi(q_1)} \diamond \boldsymbol{g} \tag{2.21}$$

with the square-root taken entrywise, and $\boldsymbol{g} \sim \mathcal{N}(0, I_N)$. Then

$$\vec{R}(\boldsymbol{n}^{\underline{\ell}}, \boldsymbol{n}^{\underline{\ell}}) \simeq \Phi(q_1+\delta) = \Phi(q_{\underline{\ell}+1}^\delta). \tag{2.22}$$

The point $\boldsymbol{n}^{\underline{\ell}}$ will be the "root" of our IAMP algorithm.[2]

Moreover we set $\overline{\ell} = \max\{\ell \in \mathbb{Z}_+ : q_\ell^\delta \leq 1 - 2\delta\}$. We also define for $s \in \mathcal{S}$ and $\underline{\ell} \leq \ell \leq \overline{\ell}$ the constants

$$u_{\ell,s}^\delta = \sqrt{\frac{\Phi_s(q_{\ell+1}^\delta) - \Phi_s(q_\ell^\delta)}{\xi^s(\Phi(q_{\ell+1}^\delta)) - \xi^s(\Phi(q_\ell^\delta))}}. \tag{2.23}$$

[2] If $\vec{h} = 0$, one takes $\underline{\ell} = q_1 = 0$, $n_i^1 = \sqrt{\Phi_{s(i)}(\delta)}\boldsymbol{g}_i$, and proceeds identically.

Set $z^{\underline{\ell}} = w^{\underline{\ell}} - h$. We will define $(z^\ell)_{\ell \geq \underline{\ell}+1}$ via

$$z^{\ell+1} = \nabla \widetilde{H}_N(f_\ell(z^{\underline{\ell}}, \dots, z^\ell)) - \sum_{j=0}^{\ell} d_{\ell,j} \diamond f_{j-1}(z^{\underline{\ell}}, \dots, z^{j-1}). \tag{2.24}$$

The Onsager coefficients $d_{\ell,j}$ are given by (2.7) and will not appear explicitly in any calculations until Sect. 3.2. Note that formally, they may depend on the first $\underline{\ell}$ iteratates, since (2.24) is a continuation of the same AMP iteration. To complete the definition of the iteration (2.24), for $s(i) = s$ and $\ell \geq \underline{\ell}$ we set

$$f_{\ell,s}(z_i^{\underline{\ell}}, \dots, z_i^\ell) = n_i^\ell, \tag{2.25}$$

where

$$n^{\ell+1} = n^\ell + u_\ell^\delta \diamond \left(z^{\ell+1} - z^\ell\right). \tag{2.26}$$

The algorithm $\mathcal{A}$ outputs

$$\mathcal{A}(H_N) = \vec{R}(n^{\overline{\ell}}, n^{\overline{\ell}})^{-1/2} \diamond n^{\overline{\ell}} \in \mathcal{B}_N \tag{2.27}$$

where the power $-1/2$ is taken entry-wise. We show in (2.32) below that

$$\lim_{\underline{\ell} \to \infty} \operatorname{p-lim}_{N \to \infty} \|n^{\overline{\ell}} - \mathcal{A}(H_N)\|_N = 0.$$

Hence we will often not distinguish between the two and just consider $n^{\overline{\ell}}$ to be the output. This makes essentially no difference by virtue of Proposition 1.6.

The state evolution limits of z^ℓ and n^ℓ are described by time-changed Brownian motions with total variance $\Phi_s(q_\ell^\delta)$ in species s after iteration ℓ. This is made precise below.

Lemma 2.6 *Fix $s \in \mathcal{S}$. The sequences $(Z^\delta_{\underline{\ell},s}, Z^\delta_{\underline{\ell}+1,s}, \dots)$ and $(N^\delta_{\underline{\ell},s}, N^\delta_{\underline{\ell}+1,s}, \dots)$ are Gaussian processes satisfying*

$$\mathbb{E}[(Z^\delta_{\ell+1,s} - Z^\delta_{\ell,s}) Z^\delta_{j,s}] = 0, \quad \text{for all } \underline{\ell} + 1 \leq j \leq \ell \tag{2.28}$$

$$\mathbb{E}\left[(Z^\delta_{\ell+1,s} - Z^\delta_{\ell,s})^2\right] = \xi^s(\Phi(q^\delta_{\ell+1})) - \xi^s(\Phi(q^\delta_\ell)) \tag{2.29}$$

$$\mathbb{E}[Z^\delta_{\ell,s} Z^\delta_{j,s}] = \xi^s(\Phi(q^\delta_{j\wedge\ell})) \tag{2.30}$$

$$\mathbb{E}[N^\delta_{\ell,s} N^\delta_{j,s}] = \Phi_s(q^\delta_{(j\wedge\ell)+1}). \tag{2.31}$$

Proof The fact that these sequences are Gaussian processes is a general fact about state evolution (the external Gaussian g is permitted in Theorem 2). We proceed by induction on $\ell \geq \underline{\ell}$. The proof is similar to [24, Sect. 8] so we give only the main points (in fact (2.21) simplifies the corresponding construction therein, which avoided the use of external Gaussian noise). We will make liberal use of (2.8) to connect asymptotic overlaps before and after applying $\nabla H_N(\cdot)$.

For base cases, the $\underline{\ell}$ case of (2.30) is immediate from (2.16). The base case of (2.31) follows from (2.22), and thus the $\underline{\ell}+1$ case of (2.30). The main computation for the base case is

$$\begin{aligned}
\mathbb{E}\left[\left(Z^\delta_{\underline{\ell}+1,s} - Z^\delta_{\underline{\ell},s}\right) Z^\delta_{\underline{\ell},s}\right] &= \xi^s\left(\{\mathbb{E}[N^\delta_{\underline{\ell},s} M_s^{\underline{\ell}-1}]\}_{s\in\mathcal{S}}\right) - \xi^s\left(\{\mathbb{E}[M_s^{\underline{\ell}-1} M_s^{\underline{\ell}-1}]\}_{s\in\mathcal{S}}\right) \\
&= \xi^s(\Phi(q_1)) - \xi^s(\Phi(q_1)) \\
&= 0.
\end{aligned}$$

Here we used the general AMP statement of Theorem 2 to say that

$$\mathbb{E}[N^{\delta}_{\underline{\ell},s} M_s^{\underline{\ell}-1}] = \mathbb{E}[M_s^{\underline{\ell}-1} M_s^{\underline{\ell}-1}] = \Phi_s(q_1).$$

For inductive steps, we always have by state evolution

$$\mathbb{E}[Z^{\delta}_{\ell+1,s} Z^{\delta}_{j+1,s}] \simeq \xi^s\left(\vec{R}(\boldsymbol{n}^{\ell}, \boldsymbol{n}^{j})\right).$$

It follows by the inductive hypothesis of (2.28) that for $j \le \ell$,

$$\begin{aligned} R_s(\boldsymbol{n}^{\ell}, \boldsymbol{n}^{j}) &= R_s(\boldsymbol{n}^{\underline{\ell}}, \boldsymbol{n}^{\underline{\ell}}) + \sum_{k=\underline{\ell}}^{j-1} (u_k^{\delta})^2 R_s(z^{k+1} - z^k, z^{k+1} - z^k) \\ &= R_s(\boldsymbol{n}^{\underline{\ell}}, \boldsymbol{n}^{\underline{\ell}}) + \sum_{k=\underline{\ell}}^{j-1} (u_k^{\delta})^2 \left(\xi^s(\Phi(q_{k+1}^{\delta})) - \xi^s(\Phi(q_k^{\delta}))\right) \\ &= \Phi_s(q_1) + \sum_{k=\underline{\ell}}^{j-1} \left(\Phi_s(q_{k+1}^{\delta}) - \Phi_s(q_k^{\delta})\right) \\ &= \Phi_s(q_j^{\delta}). \end{aligned}$$

Plugging into the above yields that for $j \le \ell$,

$$\mathbb{E}[Z^{\delta}_{\ell+1,s} Z^{\delta}_{j+1,s}] = \xi^s(\Phi(q_j^{\delta})).$$

This depends only on $\min(j, \ell)$, so (2.28) follows. The others are proved by similar computations. □

Equation (2.31) implies that $\vec{R}(\boldsymbol{n}_{\ell}^{\delta}, \boldsymbol{n}_{j}^{\delta}) \simeq \Phi(q^{\delta}_{(\ell \wedge j)+1})$, which exactly corresponds to the previous sections of the paper. In particular it implies that the final iterate $\boldsymbol{n}_{\overline{\ell}}^{\delta}$ satisfies

$$(1 - O(\delta)) \cdot \vec{1} \preceq \vec{R}(\boldsymbol{n}_{\overline{\ell}}^{\delta}, \boldsymbol{n}_{\overline{\ell}}^{\delta}) \preceq \vec{1} \tag{2.32}$$

so the rounding step (2.27) causes only an $O(\delta)$ change in the Hamiltonian value. Finally we compute in Lemma 2.7 the energy gain from the second phase, which matches the second term in (1.8).

Lemma 2.7

$$\lim_{\underline{\ell} \to \infty} \operatorname*{p-lim}_{N \to \infty} \frac{H_N(\boldsymbol{n}^{\overline{\ell}}) - H_N\left(\boldsymbol{n}^{\underline{\ell}}\right)}{N} = \sum_{s \in \mathcal{S}} \lambda_s \int_{q_1}^{1} \sqrt{\Phi_s'(t)(\xi^s \circ \Phi)'(t)} \; dt \tag{2.33}$$

Proof Observe that $\langle h, \boldsymbol{n}^{\overline{\ell}} - \boldsymbol{n}^{\underline{\ell}} \rangle_N \simeq 0$ because the values $(N^{\delta}_{\ell,s})_{\ell \ge \underline{\ell}}$ form a martingale sequence for each $s \in \mathcal{S}$. Therefore it suffices to find the in-probability limit of $\frac{\widetilde{H}_N(\boldsymbol{n}^{\overline{\ell}}) - \widetilde{H}_N(\boldsymbol{n}^{\underline{\ell}})}{N}$. We write

$$\frac{\widetilde{H}_N(\boldsymbol{n}^{\overline{\ell}}) - \widetilde{H}_N(\boldsymbol{n}^{\underline{\ell}})}{N} = \sum_{\ell=\underline{\ell}}^{\overline{\ell}-1} \frac{\widetilde{H}_N(\boldsymbol{n}^{\ell+1}) - \widetilde{H}_N(\boldsymbol{n}^{\ell})}{N}$$

and use a Taylor series approximation for each term. In particular for $F \in C^3(\mathbb{R};\mathbb{R})$, applying Taylor's approximation theorem twice yields

$$\begin{aligned} F(1) - F(0) &= F'(0) + \frac{1}{2}F''(0) + O\big(\sup_{a\in[0,1]} |F'''(a)|\big) \\ &= F'(0) + \frac{1}{2}(F'(1) - F'(0)) + O\big(\sup_{a\in[0,1]} |F'''(a)|\big) \\ &= \frac{1}{2}(F'(1) + F'(0)) + O\big(\sup_{a\in[0,1]} |F'''(a)|\big). \end{aligned}$$

Assuming $\sup_\ell \left\|\boldsymbol{n}^\ell\right\|_N \le 1$, which holds with probability $1 - o_N(1)$ by state evolution and the definition of $\overline{\ell}$, we apply this estimate with

$$F(a) = \frac{1}{N}\widetilde{H}_N\left((1-a)\boldsymbol{n}^\ell + a\boldsymbol{n}^{\ell+1}\right).$$

The result is:

$$\begin{aligned} &\frac{1}{N}\left|\widetilde{H}_N(\boldsymbol{n}^{\ell+1}) - \widetilde{H}_N(\boldsymbol{n}^\ell) - \frac{1}{2}\left\langle \nabla\widetilde{H}_N(\boldsymbol{n}^\ell) + \nabla\widetilde{H}_N(\boldsymbol{n}^{\ell+1}), \boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell\right\rangle\right| \\ &\quad \le O\left(\underline{C}\|\boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell\|_N^3\right); \\ &\underline{C}N^{-1/2} = \sup_{\|\boldsymbol{\sigma}\|\le\sqrt{N}} \left\|\nabla^3\widetilde{H}_N(\boldsymbol{\sigma})\right\|_{\mathrm{op}}. \end{aligned}$$

Proposition 1.6 implies that for deterministic constants c, C,

$$\mathbb{P}[\underline{C} \le C] \ge 1 - e^{-cN}.$$

On the other hand for each $\underline{\ell} \le \ell \le \overline{\ell} - 1$ we have

$$\begin{aligned} \operatorname*{p\text{-}lim}_{N\to\infty} \|\boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell\|_N &= \sqrt{\sum_{s\in\mathcal{S}} \lambda_s R_s(\boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell, \boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell)} \\ &= \sqrt{\sum_{s\in\mathcal{S}} \lambda_s\big(\Phi_s(q^\delta_{\ell+2}) - \Phi_s(q^\delta_{\ell+1})\big)} \\ &= \sqrt{\delta}. \end{aligned}$$

Summing and noting that $\overline{\ell} - \underline{\ell} \le \delta^{-1}$ yields the high-probability estimate

$$\begin{aligned} &\sum_{\ell=\underline{\ell}}^{\overline{\ell}-1} \frac{1}{N}\left|\widetilde{H}_N(\boldsymbol{n}^{\ell+1}) - \widetilde{H}_N(\boldsymbol{n}^\ell) - \frac{1}{2}\left\langle \nabla\widetilde{H}_N(\boldsymbol{n}^\ell) + \nabla\widetilde{H}_N(\boldsymbol{n}^{\ell+1}), \boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell\right\rangle\right| \\ &\quad \le \sum_{\ell=\underline{\ell}}^{\overline{\ell}-1} \|\boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell\|_N^3 \le O(\sqrt{\delta}). \end{aligned}$$

So, this term vanishes as $\delta \to 0$. It remains to prove

$$\lim_{\delta\to 0} \operatorname*{p\text{-}lim}_{N\to\infty} \sum_{\ell=\underline{\ell}}^{\overline{\ell}-1} \left\langle \nabla\widetilde{H}_N(\boldsymbol{n}^\ell) + \nabla\widetilde{H}_N(\boldsymbol{n}^{\ell+1}), \boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell\right\rangle_N \stackrel{?}{=} 2\sum_{s\in\mathcal{S}} \lambda_s \int_{q_1}^1 \sqrt{\Phi_s'(t)(\xi^s \circ \Phi)'(t)}\mathrm{d}t.$$

To establish this it suffices to show for each species $s \in \mathbb{S}$ the equality

$$\lim_{\delta\to 0} \operatorname*{p-lim}_{N\to\infty} \sum_{\ell=\underline{\ell}}^{\overline{\ell}-1} R_s\left(\nabla \widetilde{H}_N(\boldsymbol{n}^\ell) + \nabla \widetilde{H}_N(\boldsymbol{n}^{\ell+1}), \boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell\right) \stackrel{?}{=} 2\int_{q_1}^1 \sqrt{\Phi_s'(t)(\xi^s \circ \Phi)'(t)}\mathrm{d}t. \tag{2.34}$$

Observe by (2.24) that

$$\nabla \widetilde{H}_N(\boldsymbol{n}^\ell) = \boldsymbol{z}^{\ell+1} + \sum_{j=0}^{\ell} d_{\ell,j} \diamond \boldsymbol{n}^{j-1}. \tag{2.35}$$

Passing to the limiting Gaussian process $(Z_k^\delta)_{k\in\mathbb{Z}^+}$ via state evolution,

$$\begin{aligned}
\operatorname*{p-lim}_{N\to\infty} R\left(\nabla \widetilde{H}_N(\boldsymbol{n}^\ell), \boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell\right)_s &= \mathbb{E}\left[Z^\delta_{\ell+1,s}(N^\delta_{\ell+1,s} - N^\delta_{\ell,s})\right] \\
&\quad + \sum_{j=0}^{\ell} d_{\ell,j,s}\mathbb{E}\left[N^\delta_{j-1,s}(N^\delta_{\ell+1,s} - N^\delta_{\ell,s})\right], \\
\operatorname*{p-lim}_{N\to\infty} R\left(\nabla \widetilde{H}_N(\boldsymbol{n}^{\ell+1}), \boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell\right)_s &= \mathbb{E}\left[Z^\delta_{\ell+2,s}(N^\delta_{\ell+1,s} - N^\delta_{\ell,s})\right] \\
&\quad + \sum_{j=0}^{\ell+1} d_{\ell+1,j,s}\mathbb{E}\left[N^\delta_{j-1}(N^\delta_{\ell+1,s} - N^\delta_{\ell,s})\right].
\end{aligned}$$

As $(N_k^\delta)_{k\geq\mathbb{Z}^+}$ is a martingale process, it follows that all right-most expectations vanish. Similarly it holds that

$$\begin{aligned}
\mathbb{E}[Z^\delta_{\ell+2}(N^\delta_{\ell+1} - N^\delta_\ell)] &= \mathbb{E}[Z^\delta_{\ell+1}(N^\delta_{\ell+1} - N^\delta_\ell)] \\
\mathbb{E}[Z^\delta_{\ell}(N^\delta_{\ell+1} - N^\delta_\ell)] &= 0.
\end{aligned}$$

We conclude that

$$\begin{aligned}
&\operatorname*{p-lim}_{N\to\infty} R\left(\nabla \widetilde{H}_N(\boldsymbol{n}^\ell) + \nabla \widetilde{H}_N(\boldsymbol{n}^{\ell+1}), \boldsymbol{n}^{\ell+1} - \boldsymbol{n}^\ell\right)_s \\
&= 2\,\mathbb{E}[(Z^\delta_{\ell+1,s} - Z^\delta_{\ell,s})(N^\delta_{\ell+1,s} - N^\delta_{\ell,s})] \\
&= 2\,\mathbb{E}[u^\delta_{\ell,s}(Z^\delta_{\ell,s})(Z^\delta_{\ell+1,s} - Z^\delta_{\ell,s})^2] \\
&= 2\,\mathbb{E}[u^\delta_{\ell,s}(Z^\delta_{\ell,s})] \cdot \mathbb{E}[(Z^\delta_{\ell+1,s} - Z^\delta_{\ell,s})^2] \\
&= 2\sqrt{\left(\Phi_s(q^\delta_{\ell+1,s}) - \Phi_s(q^\delta_{\ell,s})\right) \cdot \left(\xi^s(\Phi(q^\delta_{\ell+1})) - \xi^s(\Phi(q^\delta_\ell))\right)}.
\end{aligned}$$

In the second-to-last step we used independence of $Z^\delta_{\ell,s}$ increments, which follows from Lemma 2.6, while the last step used (2.23) and (2.29). Combining with [16, Lemma 3.7] on discrete approximation of the integral in $\mathbb{A}$ implies (2.34). □

Proof of Theorem 1 We take $\mathcal{A}$ as in (2.27) for $\underline{\ell}$ a large constant depending on $(\varepsilon, \xi, h, \lambda)$. First,

$$\mathbb{P}[H_N(\mathcal{A}(H_N))/N \geq \mathsf{ALG} - \varepsilon/2] \geq 1 - o_N(1) \tag{2.36}$$

follows from combining Lemmas 2.5, 2.7 and the fact that (recall (2.32))

$$H_N(\mathcal{A}(H_N))/N \simeq H_N(\boldsymbol{n}^{\underline{\ell}})/N + o_{\mathbb{P}}(1).$$

Next, let $K_N \subseteq \mathcal{H}_N$ be as in Proposition 1.6. We recall that $\mathbb{P}[H_N \in K_N] \geq 1 - e^{-cN}$. Exactly as in [15, Theorem 10] it follows that there is a $C(\varepsilon)$-Lipschitz function $\widetilde{\mathcal{A}} : \mathcal{H}_N \to \mathbb{R}$ such that $\widetilde{\mathcal{A}}$ and $\mathcal{A}$ agree on K_N. Moreover (1.6) and concentration of measure on Gaussian space imply that $H_N(\widetilde{\mathcal{A}}(H_N))$ is $O(N^{1/2})$-sub-Gaussian. In light of (2.36) and since $\mathbb{P}[\widetilde{\mathcal{A}}(H_N) = \mathcal{A}(H_N)] \geq \mathbb{P}[H_N \in K_N] \geq 1 - e^{-cN}$, we deduce that

$$\mathbb{P}[H_N(\mathcal{A}(H_N))/N \geq \mathsf{ALG} - \varepsilon] \geq 1 - e^{-cN}.$$

This concludes the proof. □

3 Extensions

3.1 Signed AMP

In our companion paper [17], we show that strictly super-solvable models have w.h.p. exactly 2^r critical points, indexed by sign patterns $\vec{\Delta} \in \{\pm 1\}^r$ with the following physical meaning. Consider first the extreme case of a linear Hamiltonian, with external field $\boldsymbol{h} = \vec{h} \diamond \mathbf{1}$ where all entries of $\vec{h}$ are nonzero and **no other interactions**. This model clearly has 2^r critical points, which are the products of the maxima and minima in the spheres $\{\|\boldsymbol{x}_s\|_2^2 = \lambda_s N\}$ corresponding to each species $s \in \mathcal{S}$, and the signs $\vec{\Delta}$ record whether the critical point is a maximum or minimum in each species. As explained in [17, Sect. 6.6], if a strictly super-solvable H_N is gradually deformed to a linear function (staying inside the strictly super-solvable phase), the critical points move stably, and over this process their Hessian eigenvalues do not cross zero. Thus, each critical point of H_N can also be associated with a sign pattern $\vec{\Delta}$.

We now show that the root-finding algorithm defined in Sect. 2.3 can be generalized to find all 2^r critical points in a strictly super-solvable model. More precisely, it finds 2^r approximate critical points, one in a neighborhood of each exact critical point of the model, from which the exact critical points can be computed by Newton's method (see Remark 3.2). For general models, it finds 2^r approximate critical points on the product of spheres with self-overlap $\Phi(q_1)$. The restriction of H_N to this set, considered as a spin glass in its own right (see [16, Remark 1.2]) is a solvable model.

Fixing $\vec{\Delta} \in \{\pm 1\}^r$, the analogous iteration to (2.10) is:

$$\begin{aligned} \boldsymbol{w}^{k+1} &= \nabla H_N(\boldsymbol{m}^k) - \vec{b}_k \diamond \boldsymbol{m}^{k-1} \\ &= \boldsymbol{h} + \nabla \widetilde{H}_N(\boldsymbol{m}^k) - \vec{b}_k(\vec{\Delta}) \diamond \boldsymbol{m}^{k-1}; \\ \boldsymbol{m}^k &= \vec{\Delta} \odot \vec{a} \diamond \boldsymbol{w}^k \\ b_{k,s}(\vec{\Delta}) &\equiv \sum_{s' \in \mathcal{S}} \Delta_{s'} a_{s'} \partial_{s'} \xi^s \big(\vec{R}(\boldsymbol{m}^k, \boldsymbol{m}^{k-1}) \big). \end{aligned} \tag{3.1}$$

The change of sign does not affect the proofs or statements of Lemmas 2.3, 2.4. Indeed a_s^2 only changes to $\Delta_s^2 a_s^2$ in the former proof which is no change at all. The generalization of Lemma 2.5 is as follows.

Lemma 3.1

$$\lim_{k \to \infty} \operatorname*{p-lim}_{N \to \infty} \frac{H_N(\boldsymbol{m}^k)}{N} = \sum_{s \in \mathcal{S}} \lambda_s \Delta_s \sqrt{\Phi_s(q_1) \left(h_s^2 + \xi^s(\Phi(q_1)) \right)}.$$

Proof The proof is similar to Lemma 2.5. The main calculation now becomes:

$$\begin{aligned}\int_0^1 \langle \boldsymbol{m}^k, \nabla \widetilde{H}_N(t\boldsymbol{m}^k)\rangle_N \mathrm{d}t &\simeq \sum_{s\in\mathcal{S}} \lambda_s \int_0^1 R_s(\boldsymbol{m}^k, \nabla \widetilde{H}_N(t\boldsymbol{m}^k))\mathrm{d}t \\ &\simeq \sum_{s\in\mathcal{S}} \lambda_s \int_0^1 \left(\Delta_s a_s \xi^s(t\vec{R}^k) + t R_s^k \sum_{s'\in\mathcal{S}} \Delta_{s'} a_{s'} \partial_{s'} \xi^s(\vec{R}^k)\right) \mathrm{d}t \\ &\overset{(2.20)}{=} \sum_{s\in\mathcal{S}} \lambda_s \int_0^1 \left(\Delta_s a_s \xi^s(t\vec{R}^k) + t \Delta_s a_s \sum_{s'\in\mathcal{S}} R_{s'}^k \partial_{s'} \xi^s(\vec{R}^k)\right) \mathrm{d}t \\ &= \sum_{s\in\mathcal{S}} \lambda_s \Delta_s a_s \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\left(t\, \xi^s(t\ \vec{R}^k)\right) \mathrm{d}t \\ &= \sum_{s\in\mathcal{S}} \lambda_s \Delta_s a_s \xi^s(\vec{R}^k).\end{aligned}$$

Moreover the external field $\boldsymbol{h}$ now contributes energy

$$\langle \boldsymbol{h}, \boldsymbol{m}^k\rangle_N \simeq \sum_{s\in\mathcal{S}} \lambda_s h_s \mathbb{E}[M_s^k] = \sum_{s\in\mathcal{S}} \lambda_s \Delta_s a_s h_s^2.$$

Combining gives the desired statement. □

Remark 3.1 One can sign the IAMP phase as well by redefining (2.26) to

$$\boldsymbol{n}^{\ell+1}(\vec{\Delta}) = \boldsymbol{n}^{\ell}(\vec{\Delta}) + \vec{\Delta} \odot u_\ell^\delta \diamond (z^{\ell+1} - z^\ell). \tag{3.2}$$

The resulting output $\boldsymbol{n}^{\overline{\ell}}(\vec{\Delta})$ then achieves asymptotic energy (recall (1.8))

$$\begin{aligned}&\lim_{\underline{\ell}\to\infty} \operatorname*{p\text{-}lim}_{N\to\infty} \frac{H_N\left(\boldsymbol{n}^{\overline{\ell}}(\vec{\Delta})\right)}{N} \\ &= \sum_{s\in\mathcal{S}} \lambda_s \Delta_s \left[\sqrt{\Phi_s(q_1)(\xi^s(\Phi(q_1)) + h_s^2)} + \int_{q_1}^1 \sqrt{\Phi_s'(q)(\xi^s \circ \Phi)'(q)}\, \mathrm{d}q\right].\end{aligned} \tag{3.3}$$

However it is unclear whether $\boldsymbol{n}^{\overline{\ell}}(\vec{\Delta})$ can be made to obey any notable properties. We will show that the signed outputs $\boldsymbol{m}^k(\vec{\Delta})$ of the first phase above are approximate critical points for H_N (and in [17] that all near-critical points are close to one of them). By contrast, for the output of signed IAMP to be a critical point, Φ must satisfy a signed version of the tree-descending ODE (2.3) in which the function $(\xi^s \circ \Phi)'(q)$ is replaced by

$$\sum_{s'\in\mathcal{S}} \Delta_{s'} \partial_{s'} \xi^s(\Phi(q)) \Phi_s'(q).$$

Since this quantity appears inside a square root in (2.3), it is unclear when to expect solutions to exist. Furthermore the proof in [16] of well-posedness relies on positivity of coefficients (via Perron-Frobenius theory) and does not seem to generalize. Additionally, a solution would not seem to correspond to a maximizer of any variational problem as in (1.8). As a result we do not know how to prove a solution exists in the signed case. However if one takes *as given* a smooth function Φ satisfying the signed tree-descending ODE, the iteration (3.2) starting from signed initialization $\boldsymbol{n}^{\underline{\ell}}(\vec{\Delta}) = \boldsymbol{m}^{\underline{\ell}}(\vec{\Delta}) + \sqrt{\Phi(q_1+\delta) - \Phi(q_1)} \diamond \boldsymbol{g}$ would produce an approximate critical point $\boldsymbol{n}^{\overline{\ell}}(\vec{\Delta})$ which still satisfies (3.3).

3.2 Gradient Computation and Connection to E_∞

We now compute the gradient of the outputs, showing that $\boldsymbol{m}^{\underline{\ell}}(\vec{\Delta})$ and $\boldsymbol{n}^{\ell}$ ($\underline{\ell} \le \ell \le \overline{\ell}$) are approximate critical points for the restriction of H_N to the products of r spheres with suitable radii passing through them. For $\boldsymbol{\sigma}$ to be an approximate critical point means precisely that there exist coefficients $\vec{A} \in \mathbb{R}^r$ such that

$$\|\nabla H_N(\boldsymbol{\sigma}) - \vec{A} \diamond \boldsymbol{\sigma}\|_N \simeq 0.$$

In our case, these coefficients will be given as follows. If $\vec{1}$ is strictly sub-solvable (so $q_1 < 1$), define $\vec{A}(q)$ for $q \in [q_1, 1]$ by

$$A_s(q) \equiv f_s(q)^{-1} + \sum_{s' \in \mathcal{S}} f_{s'}(q) \partial_{s'} \xi^s\big(\Phi(q)\big), \tag{3.4}$$

$$f_s(q) \equiv \sqrt{\frac{\Phi_s'(q)}{(\xi^s \circ \Phi)'(q)}}. \tag{3.5}$$

Further define for $\vec{\Delta} \in \{-1, 1\}^r$

$$A_s(q_1; \vec{\Delta}) \equiv \Delta_s \sqrt{\frac{\xi^s(\Phi(q_1)) + h_s^2}{\Phi_s(q_1)}} + \sum_{s' \in \mathcal{S}} \Delta_{s'} \partial_{s'} \xi^s\big(\Phi(q_1)\big) \sqrt{\frac{\Phi_{s'}(q_1)}{\xi^{s'}(\Phi(q_1)) + h_{s'}^2}}. \tag{3.6}$$

Note that, by (2.2), this is consistent with the definition of $\vec{A}(q_1)$ above, in the sense that $\vec{A}(q_1; \vec{1}) = \vec{A}(q_1)$. We take this to be the definition of $\vec{A}(q_1)$ if $\vec{1}$ is super-solvable (and $q_1 = 1$).

Proposition 3.2 *If Φ is a pseudo-maximizer for $\mathbb{A}$ (recall Definition 2.1) then for any $\vec{\Delta} \in \{\pm 1\}^r$,*

$$\lim_{\underline{\ell} \to \infty} \operatorname*{p-lim}_{N \to \infty} \|\nabla H_N(\boldsymbol{m}^{\underline{\ell}}(\vec{\Delta})) - \vec{A}(q_1, \vec{\Delta}) \diamond \boldsymbol{m}^{\underline{\ell}}(\vec{\Delta})\|_N = 0. \tag{3.7}$$

Proof Recall from Lemma 2.3 (which holds without modification for general $\vec{\Delta}$) that

$$\lim_{\underline{\ell} \to \infty} \operatorname*{p-lim}_{N \to \infty} \|\boldsymbol{m}^{\underline{\ell}+1}(\vec{\Delta}) - \boldsymbol{m}^{\underline{\ell}}(\vec{\Delta})\|_N = 0. \tag{3.8}$$

Thus rearranging (3.1) yields

$$\lim_{k \to \infty} \operatorname*{p-lim}_{N \to \infty} \|\nabla H_N(\boldsymbol{m}^{\underline{\ell}}(\vec{\Delta})) - (\vec{\Delta} \odot \vec{a}^{-1} + \vec{b}_k(\vec{\Delta})) \diamond \boldsymbol{m}^{\underline{\ell}}(\vec{\Delta})\|_N = 0.$$

Since $\lim_{\underline{\ell} \to \infty} \big(\Delta_s a_s^{-1} + b_{\underline{\ell},s}(\vec{\Delta})\big) = A_s(\vec{\Delta})$ by (3.6), the result follows. □

Remark 3.2 In [17, Theorems 1.5 and 1.6], we show that when ξ is strictly super-solvable, H_N has exactly 2^r critical points $\{\boldsymbol{x}(\vec{\Delta})\}_{\vec{\Delta} \in \{-1,1\}^r}$. Moreover all ε-approximate critical points with Riemannian gradient $\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x})\| \le \varepsilon\sqrt{N}$ are within $o_\varepsilon(\sqrt{N})$ of some $\boldsymbol{x}(\vec{\Delta})$. It follows from Proposition 3.2 that each $\boldsymbol{m}^{\underline{\ell}}(\vec{\Delta})$ is an ε-approximate critical point for large enough $\underline{\ell} = \underline{\ell}(\xi, \varepsilon)$. In fact the preceding gradient computation shows that the values $\vec{\Delta}$ agree, implying that $\|\boldsymbol{m}^{\underline{\ell}}(\vec{\Delta}) - \boldsymbol{x}(\vec{\Delta})\|_N \le o_{\underline{\ell} \to \infty}(1)$ (compare with [17, Definition 5, Eq. (1.15)]). Moreover by [17, Theorem 1.6] each Riemannian Hessian $\nabla_{\mathrm{sp}}^2 H_N(\boldsymbol{x}(\vec{\Delta}))$ has condition number at least $1/C(\xi)$. It follows that each critical point $\boldsymbol{x}(\vec{\Delta})$ can be efficiently computed to arbitrary accuracy by applying Newton's method from $\boldsymbol{m}^{\underline{\ell}}(\vec{\Delta})$ for a large enough $\underline{\ell} = \underline{\ell}(\xi)$. (By contrast, the convergence of $\boldsymbol{m}^{\underline{\ell}}(\vec{\Delta})$ itself to $\boldsymbol{x}(\vec{\Delta})$ is only in the careful double-limit sense $\lim_{\underline{\ell} \to \infty} \lim_{N \to \infty}$.)

Proposition 3.3 *If Φ is a pseudo-maximizer for $\mathbb{A}$, then for any $\underline{\ell}$-indexed sequence $(q_*, \ell) = \big((q_*, \ell)_{\underline{\ell} \geq 1}\big)$ such that $q_* \in [q_1, 1]$, $\underline{\ell} \leq \ell \leq \overline{\ell}$ and $\lim_{\underline{\ell} \to \infty} |q_* - q_\ell^\delta| = 0$, we have*

$$\lim_{\underline{\ell} \to \infty} \operatorname*{p-lim}_{N \to \infty} \left\| \nabla H_N(\boldsymbol{n}^\ell) - \vec{A}(q_*) \diamond \boldsymbol{n}^\ell \right\|_N = 0.$$

Proof For notational convenience we assume $(q_*, \ell) = (1, \overline{\ell})$; the proof is identical in general. Recall the rearrangement (2.35):

$$\nabla \widetilde{H}_N(\boldsymbol{n}^{\overline{\ell}}) = \boldsymbol{z}^{\overline{\ell}+1} + \sum_{j=0}^{\overline{\ell}} d_{\overline{\ell},j} \diamond \boldsymbol{n}^{j-1}. \tag{3.9}$$

So far we did not have to compute $d_{\overline{\ell},j}$. We do this now, focusing on the IAMP phase. Recalling (2.25), the IAMP iteration used non-linearity

$$\begin{aligned} \boldsymbol{f}_{\overline{\ell}} = \boldsymbol{n}^{\overline{\ell}} &= n^{\underline{\ell}} + \sum_{j=\underline{\ell}}^{\overline{\ell}-1} (\boldsymbol{n}^{j+1} - \boldsymbol{n}^j) \\ &= \vec{a} \diamond \boldsymbol{z}^{\underline{\ell}} + \sum_{j=\underline{\ell}}^{\overline{\ell}-1} \boldsymbol{u}_j^\delta \diamond (\boldsymbol{z}^{j+1} - \boldsymbol{z}^j) \\ &= (\vec{a} - \boldsymbol{u}_{\underline{\ell}}^\delta) \diamond \boldsymbol{z}^{\underline{\ell}} + \boldsymbol{u}_{\overline{\ell}-1}^\delta \diamond \boldsymbol{z}^{\overline{\ell}} - \sum_{j=\underline{\ell}+1}^{\overline{\ell}-1} (\boldsymbol{u}_j^\delta - \boldsymbol{u}_{j-1}^\delta) \diamond \boldsymbol{z}^j. \end{aligned}$$

Using the formula (2.7) we find

$$d_{\overline{\ell},j,s} \approx \begin{cases} \sum_{s' \in \mathbb{S}} \partial_{s'} \xi^s(\Phi(1))\, u_{\overline{\ell},s'}^\delta, & j = \overline{\ell}; \\ -\sum_{s' \in \mathbb{S}} \partial_{s'} \xi^s(\Phi(q_{j-1}^\delta))\, (u_{j,s'}^\delta - u_{j-1,s'}^\delta), & \underline{\ell} < j < \overline{\ell}; \\ \sum_{s' \in \mathbb{S}} \partial_{s'} \xi^s(\Phi(q_1))\, (a_{s'} - u_{\underline{\ell},s'}^\delta), & j = \underline{\ell}. \end{cases}$$

Note that since $\Phi \in C^2([q_1, 1])$ we have the uniform-in-q_j^δ approximations (recall (3.5)):

$$\begin{aligned} u_{\ell,s}^\delta &\approx f_s(q_j^\delta), \\ \frac{u_{\ell,s}^\delta - u_{\ell-1,s}^\delta}{\delta} &\approx \frac{\mathrm{d}}{\mathrm{d}q} \sqrt{\frac{\Phi_s'(q)}{(\xi^s \circ \Phi)'(q)}} \Bigg|_{q = q_j^\delta}, \\ a_s &\approx u_{\underline{\ell},s}^\delta \approx f_s(q_{\underline{\ell}}^\delta) \approx f_s(q_1). \end{aligned} \tag{3.10}$$

Substituting into (3.9), we obtain

$$\begin{aligned} \nabla H_N(\boldsymbol{n}^{\overline{\ell}}) &= \boldsymbol{h} + \boldsymbol{z}^{\overline{\ell}+1} + \sum_{j=\underline{\ell}}^{\overline{\ell}} d_{\overline{\ell},j} \diamond \boldsymbol{n}^{j-1} \\ &= \vec{a}^{-1} \diamond \boldsymbol{m}^{\underline{\ell}} + \sum_{j=\underline{\ell}}^{\overline{\ell}} u_j^\delta \diamond (\boldsymbol{n}^{j+1} - \boldsymbol{n}^j) + \sum_{j=\underline{\ell}}^{\overline{\ell}} d_{\overline{\ell},j} \diamond \boldsymbol{n}^{j-1} \\ &\approx \Big(\vec{a}^{-1} + \sum_{j=\underline{\ell}}^{\overline{\ell}} d_{\overline{\ell},j}\Big) \diamond \boldsymbol{n}^{\underline{\ell}} + \sum_{j=\underline{\ell}}^{\overline{\ell}-1} \vec{C}_j \diamond (\boldsymbol{n}^{j+1} - \boldsymbol{n}^j); \end{aligned}$$

$$C_{j,s} \equiv a_s^{-1} + \sum_{k=j+1}^{\underline{\ell}} d_{\overline{\ell},k,s}$$
$$\overset{(3.10)}{\approx} f_s(q_j^\delta)^{-1} + \sum_{s' \in \mathcal{S}} \left(\partial_{s'} \xi^s(\Phi(1)) f_{s'}(1) - \int_{q_j^\delta}^1 \partial_{s'} \xi^s(\Phi(q))\, f'_{s'}(q)\, \mathrm{d}q \right)$$
$$\equiv \widehat{C}_s(q_j^\delta). \tag{3.11}$$

Since the increments $(\boldsymbol{n}^{j+1} - \boldsymbol{n}^j)$ are orthogonal in the state evolution sense, it easily follows that the approximation of $C_{j,s}$ by $\widehat{C}_s(q_j^\delta)$ commutes with summation, i.e.

$$\nabla H_N(\boldsymbol{n}^{\overline{\ell}}) \approx \left(\vec{a}^{-1} + \sum_{j=\underline{\ell}}^{\overline{\ell}} d_{\overline{\ell},j} \right) \diamond \boldsymbol{n}^{\underline{\ell}} + \sum_{j=\underline{\ell}}^{\overline{\ell}-1} \widehat{C}(q_j^\delta) \diamond (\boldsymbol{n}^{j+1} - \boldsymbol{n}^j)$$

Note that we manifestly have $\widehat{C}(1) = \vec{A}(1)$. We claim the function $\widehat{C}$ is constant on $[q_1, 1]$. This is equivalent to showing that for each s the function

$$F_s(q) = \frac{1}{f_s(q)} + \left(\int_{q_1}^q \sum_{s' \in \mathcal{S}} \partial_{s'} \xi^s(\Phi(t)) f'_{s'}(t)\, \mathrm{d}t \right)$$

is constant. Differentiating, it suffices to show

$$\sum_{s' \in \mathcal{S}} \partial_{s'} \xi^s(\Phi(q)) f'_{s'}(q) \overset{?}{=} f'_s(q)/f_s(q)^2. \tag{3.12}$$

Write $f'_{s'}(q) = \Psi(q)\Phi'_{s'}(q)$, where Ψ is independent of s since Φ solves the tree-descending ODE (2.3). Then using the chain rule, the left-hand side of (3.12) equals

$$\Psi(q) \sum_{s' \in \mathcal{S}} \partial_{s'} \xi^s(\Phi(q)) \cdot \Phi'_{s'}(q) = \Psi(q)(\xi^s \circ \Phi)'(q).$$

Meanwhile the right-hand side of (3.12) is

$$f'_s(q)/f_s(q)^2 = \Psi(q)\Phi'_s(q) \cdot \frac{(\xi^s \circ \Phi)'(q)}{\Phi'_s(q)} = \Psi(q)(\xi^s \circ \Phi)'(q).$$

Therefore $\widehat{C}(q) = \vec{A}(1)$ is constant as claimed. Finally it is clear that the $\boldsymbol{n}^{\underline{\ell}}$ coefficient in (3.11) approximately equals $\widehat{C}(q_1)$ and hence also $\vec{A}(1)$. Then (3.11) implies

$$\nabla H_N(\boldsymbol{n}^{\overline{\ell}}) \approx \vec{A}(1) \diamond \left(\boldsymbol{n}^{\underline{\ell}} + \sum_{j=\underline{\ell}}^{\overline{\ell}-1} (\boldsymbol{n}^{j+1} - \boldsymbol{n}^j) \right)$$
$$= \vec{A}(1) \diamond \boldsymbol{n}^{\overline{\ell}}$$

which completes the proof. □

From the point of view of [16], the fact that $\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{n}^{\overline{\ell}})\|_N \approx 0$ is to be expected. At least for $(\Phi; q_1)$ maximizing $\mathbb{A}$, if this were not true than an extra step of gradient descent would essentially suffice to reach energy strictly better than ALG, contradicting the optimality in [16, Theorem 1]. However the radial derivative computation is interesting in its own right and lets us study the spherical Hessian around an output $\boldsymbol{\sigma}$. We believe that Corollary 3.4 can be strengthened to hold with $\boldsymbol{\lambda}_1$ rather than $\boldsymbol{\lambda}_{\varepsilon N}$. This seems to require a more precise Gaussian conditioning argument around $\mathcal{A}(H_N)$ which we chose not to pursue.

Corollary 3.4 *With $\boldsymbol{\lambda}_k$ the k-th largest eigenvalue of a symmetric real matrix,*

$$\lim_{\underline{\ell}\to\infty,\varepsilon\to 0} \operatorname*{p-lim}_{N\to\infty} \boldsymbol{\lambda}_{\varepsilon N}\left(\nabla^2_{sp} H_N(\boldsymbol{n}^{\overline{\ell}})\right) = 0. \tag{3.13}$$

Proof Fixing $\vec{A} = \vec{A}(1)$, the bulk spectral measure of

$$\boldsymbol{W}(\boldsymbol{x}) = \nabla^2 H_N(\boldsymbol{x}) - \vec{A} \diamond \boldsymbol{x} \tag{3.14}$$

for deterministic $\boldsymbol{x} \in \mathcal{S}_N$ concentrates with rate function N^2 around a limiting spectral measure independent of $\boldsymbol{x}$. By union-bounding over an $\delta\sqrt{N}$-net as in [26, Proof of Lemma 3], it thus suffices to show (3.13) at a point $\boldsymbol{x} \in \mathcal{S}_N$ independent of H_N, with $\boldsymbol{W}(\boldsymbol{x})$ in place of $\nabla^2_{\mathrm{sp}} H_N(\boldsymbol{\sigma})$. This is purely a statement of random matrix theory and is shown in [17, Proposition 5.18]. □

Notably Corollary 3.4 explains the equality $\mathsf{ALG} = E_\infty$ for pure models, which we derived manually in [16]. Indeed for a pure model with $\xi = \prod_{i=1}^r x_i^{a_i}$, the energy and radial derivative are deterministically proportional:

$$\nabla_{\mathrm{rad}} H_N(\boldsymbol{x}) = -H_N(\boldsymbol{x})\vec{a} \diamond \boldsymbol{x}, \quad \forall \boldsymbol{x} \in \mathcal{B}_N.$$

It follows (using again the N^2 large deviation rate for the spectral bulk) that there is a unique energy level E_∞ at which critical points can have spherical Hessian obeying the conclusion of Corollary 3.4. This is the definition of E_∞ given in [1, 20].

3.3 Branching IAMP and Exponential Concentration

Here we modify the second stage of our IAMP algorithm (which requires $\vec{\Delta} = \vec{1}$) to use external Gaussian randomness in a small number of increment steps. This allows the construction of an ultrametric tree of outputs with large constant depth and $\exp(cN)$ breadth, with pairwise overlaps given by Φ. More precisely, for any finite ultrametric space $X = (x_1, \dots, x_M)$, $M = \exp(cN)$, of diameter at most $1 - q_1$, branching IAMP outputs $(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_M)$ with

$$\operatorname*{p-lim}_{N\to\infty} \max_{1\le i,j\le M} \|\vec{R}(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j) - \Phi\big(1 - d_X(x_i, x_j)\big)\|_\infty = 0.$$

We use an approach suggested in [3] by injecting external Gaussian noise $\boldsymbol{g}^{(i)}$ into the IAMP phase of the algorithm at depth $q_i \in (q_1, 1)$. Importantly, this gives an explicit construction of $\exp(cN)$ approximate critical points of H_N (with exponentially good probability) whenever there is an IAMP phase. A similar construction was used by one of us in [24, Sect. 4]. There the Gaussian noise was constructed artificially by preliminary iterates of AMP rather than from exogenous noise (due to the lack of a state evolution result incorporating independent Gaussian vectors). This only enabled the construction of a large constant number of outputs rather than exponentially many.

Our branching IAMP proceeds as follows. We first apply Stage I with $\vec{\Delta} = \vec{1}$ as before. We fix $q_1 < q_2 < \cdots < q_m = 1$ and let

$$\ell^\delta_{q_i} = \underline{\ell} + \left\lceil \frac{q_i - q}{\delta} \right\rceil + 1, \quad i \in [m].$$

We define $\boldsymbol{n}^\ell$ with the same recursive formula as before, unless $\ell = \ell^\delta_{q_i}$ for some $i \in [m]$. For these cases, we define $\boldsymbol{g}^{(1)}, \dots, \boldsymbol{g}^{(m)} \sim \mathcal{N}(0, \mathbf{1}_N)$ to be independent standard Gaussian

vectors. Then we set:

$$\boldsymbol{n}^{\ell+1} = \begin{cases} \boldsymbol{n}^{\ell} + \sqrt{\xi^s\big(\Phi(q^{\delta}_{\ell^{\delta}_{q_i}+1})\big) - \xi^s\big(\Phi(q^{\delta}_{\ell^{\delta}_{q_i}})\big)} \diamond \boldsymbol{g}^{(i)}, & \ell = \ell^{\delta}_{q_i} \quad \text{for some } i \in [m] \\ \boldsymbol{n}^{\ell} + u^{\delta}_{\ell} \diamond \big(\boldsymbol{z}^{\ell+1} - \boldsymbol{z}^{\ell}\big), & \text{else.} \end{cases} \tag{3.15}$$

The definition (3.15) naturally enables couplings for pairs of iterations. We say the iterations $\big(\boldsymbol{n}^{\ell,1}, \boldsymbol{n}^{\ell,2}\big)_{\ell \geq 1}$ are q_j-coupled if their associated Gaussian vectors

$$\boldsymbol{g}^{(1,1)}, \ldots, \boldsymbol{g}^{(m,1)} \sim \mathcal{N}(0, \mathbf{1}_N),$$
$$\boldsymbol{g}^{(1,2)}, \ldots, \boldsymbol{g}^{(m,2)} \sim \mathcal{N}(0, \mathbf{1}_N)$$

are coupled so that $\boldsymbol{g}^{(i,1)} = \boldsymbol{g}^{(i,2)}$ almost surely for $i < j$, and the variables are otherwise independent.

Proposition 3.5 *Let the iterations $\boldsymbol{n}^{\ell,1}, \boldsymbol{n}^{\ell,2}$ be q_j coupled as above, and let Φ be a pseudo-maximizer of $\mathbb{A}$ (recall Definition 2.1). Then*

$$\lim_{\underline{\ell} \to \infty} \operatorname*{p-lim}_{N \to \infty} \frac{H_N(\boldsymbol{n}^{\overline{\ell},a}_{\delta})}{N} = \mathbb{A}(\Phi), \quad a \in \{1, 2\}; \tag{3.16}$$

$$\lim_{\underline{\ell} \to \infty} \operatorname*{p-lim}_{N \to \infty} \Big\| \nabla H_N(\boldsymbol{n}^{\ell,a}) - \vec{A}(1) \diamond \boldsymbol{n}^{\ell,a} \Big\|_N = 0, \quad a \in \{1, 2\}; \tag{3.17}$$

$$\lim_{\underline{\ell} \to \infty} \operatorname*{p-lim}_{N \to \infty} \frac{\langle \boldsymbol{n}^{\overline{\ell},1}_{\delta}, \boldsymbol{n}^{\overline{\ell},2}_{\delta} \rangle}{N} = \Phi(q^{\delta}_j). \tag{3.18}$$

Proof The analysis uses the slightly generalized state evolution given in Theorem 2, which states that (2.8) continues to hold even in the presence of external randomness $\boldsymbol{g}^{(i)}$. Modulo this point, the calculations are essentially identical. Indeed [24] uses exactly the same calculations to analyze a slightly different formulation of branching IAMP (therein, the vectors $\boldsymbol{g}$ are defined via negatively time-indexed AMP iterates to sidestep the lack of a generalized state evolution result). We therefore give only an outline below.

The SDE description in (2.6) is unchanged if one uses the slightly added generality of Theorem 2 to incorporate the external Gaussian noise. (This Gaussian noise is scaled in (3.15) to achieve exactly the same effect as a usual iteration step.) The energy analysis of $H_N(\boldsymbol{n}^{\overline{\ell}})$ only changes on the m modified steps which has negligible effect since $\delta \to 0$ as $\underline{\ell} \to \infty$; similarly for $\nabla H_N(\boldsymbol{n}^{\overline{\ell}})$. Thus (3.16) follows by the same proof as before. The proof of (3.18) is identical to [24, Sect. 8]. □

In Proposition 3.6 we observe that concentration of measure implies Proposition 3.5 holds with exponentially high probability. Thus we can couple together $\exp(cN)$ branching IAMPs to construct a full ultrametric tree of large constant depth m and breadth $\exp(cN)$. To do this, we fix m, take $\underline{\ell}$ sufficiently large and then $\eta > 0$ sufficiently small. Then with $K = \exp(\eta N)$, we consider a complete depth m rooted tree $\mathcal{T}$, with root defined to have depth 1, such that each vertex at depths $1, \ldots, m-1$ has K children. Thus the leaf-set $L(\mathcal{T})$ is naturally indexed by $[K]^m$. For $v, v' \in L(\mathcal{T})$ we let $v \wedge v' \in \{1, 2, \ldots, m\}$ denote the height of their least common ancestor. For each non-leaf $x \in V(\mathcal{T})$, label the edge from x to its parent by an i.i.d. Gaussian vector $\boldsymbol{g}^{(x)} \sim \mathcal{N}(0, I_N)$. Then for each leaf $v \in L(\mathcal{T})$, using the m Gaussian vectors along the path from the root of $\mathcal{T}$ to v yields branching IAMP output $\boldsymbol{\sigma}^{(v)}$ for any H_N.

Proposition 3.6 *Proposition 3.5 holds with exponentially good probability in the following sense. Fix m and $q_1 < q_2 < \cdots < q_m = 1$. For any $\varepsilon > 0$, for large enough $\underline{\ell}$ there exists $\eta = \eta(\varepsilon, \underline{\ell}) > 0$ such that for N large enough, the following hold simultaneously across all $v, v' \in L(\mathcal{T})$ with probability at least $1 - \exp(-\eta N)$:*

$$\begin{aligned} |\mathbb{A}(\Phi) - H_N(\boldsymbol{n}^{\overline{\ell},v})/N| &\le \varepsilon; \\ \|\nabla H_N(\boldsymbol{n}^{\overline{\ell},v}) - \vec{A}(1) \diamond \boldsymbol{n}^{\overline{\ell},v}\|_N &\le \varepsilon; \\ \left\| \vec{R}(\boldsymbol{n}^{\overline{\ell},v}, \boldsymbol{n}^{\overline{\ell},v'}) - \Phi(q_{v\wedge v'}) \right\|_N &\le \varepsilon. \end{aligned} \tag{3.19}$$

Proof As explained in [15, Sect. 8], the map $H_N \mapsto \boldsymbol{n}^{\overline{\ell}}$ agrees with a $C(\underline{\ell})$-Lipschitz function of the coefficients $\boldsymbol{G}^{(k)}$ of H_N except with probability $1-\exp(-cN)$. The same proof applies for $H_N \mapsto \boldsymbol{n}^{\overline{\ell},v}$ as well since the external noise variables are also Gaussian. Concentration of measure on Gaussian space now ensures that the statements above hold with exponentially high probability for each fixed (v, v'). Union bounding over all such pairs for small enough η implies the result. □

In particular, the last conclusion in (3.19) shows that all $\exp(\eta N)$ constructed points have pairwise distance at least $\delta\sqrt{N}$ for $0 < \delta < 1 - q_{m-1}$. Thus for any sub-solvable model, with high probability there are exponentially many $\sqrt{N}/C(\xi)$-separated approximate critical points. This is a converse to the main result of [17], where we show that strictly super-solvable models enjoy a *strong topological trivialization* property which rules out such behavior.

Remark 3.3 An alternative to branching IAMP, which is very natural from the point of view of our companion work [16], is to slightly perturb H_N to a $(1-\eta)$-correlated function $H_N^{(\eta)}$. Concentration of measure implies that the overlap

$$\vec{R}\big(\mathcal{A}(H_N), \mathcal{A}(H_N^{(\eta)})\big)$$

concentrates exponentially around a limiting value $R_{\delta,\underline{\ell},\eta} \in \mathbb{R}^r$. We expect that taking $\eta \to 0$ with $\delta, \underline{\ell}$ in a suitable way enables $R_{\delta,\underline{\ell},\eta} \approx \Phi(q)$ for any desired $q \in [q_1, 1]$. This corresponds to the fact that $p(q)$ for $q \in [q_1, 1]$ for any $(p, \Phi; q_0)$ maximizing $\mathbb{A}$. However this approach seems more cumbersome to analyze explicitly.

Remark 3.4 The construction in this section shows the quenched existence of $\exp(\eta N)$ well-separated *approximate* critical points for strictly sub-solvable models. In [17, Theorem 5.15] we use this fact to prove the number of exact critical points is exponentially large *in expectation*. However we are unable to prove the quenched (i.e. high-probability) existence of $\exp(\eta N)$ exact critical points in strictly sub-solvable models. Showing that this is the case, or more generally identifying the quenched exponential order of the number of critical points, is an interesting direction for future work.

Acknowledgements B.H. was supported by an NSF Graduate Research Fellowship, a Siebel scholarship, NSF awards CCF-1940205 and DMS-1940092, and NSF-Simons collaboration grant DMS-2031883. M.S. was supported by an NSF graduate research fellowship, a Stanford graduate fellowship, and NSF award CCF-2006489 and was a member at the IAS while parts of this work were completed.

Funding Open Access funding provided by the MIT Libraries

Data Availability We do not analyze or generate any data. Instead, our work proceeds via a mathematical approach.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests (beyond the aforementioned funding).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Appendix A: State Evolution: Proof of Proposition 2.2

In this section we prove Proposition 2.2, following the appendix of [4]. Throughout, we denote by $\boldsymbol{G}^{(k)} \in (\mathbb{R}^N)^{\otimes k}$, $k \geq 2$ a sequence of standard Gaussian tensors. For S_k the symmetric group on k elements we also write

$$\boldsymbol{A}^{(k)} = \frac{1}{N^{(k-1)/2}} \Gamma^{(k)} \diamond \sum_{\pi \in S^k} (\boldsymbol{G}^{(k)})^{\pi} \tag{A.1}$$

for the rescaled tensors with entries

$$\boldsymbol{A}^{(k)}_{i_1,\dots,i_k} = \frac{1}{N^{(k-1)/2}} \gamma_{s(i_1),\dots,s(i_k)} \sum_{\pi \in S_k} \boldsymbol{G}^{(k)}_{i_{\pi(1)},\dots,i_{\pi(k)}}. \tag{A.2}$$

For a symmetric tensor $\boldsymbol{A}^{(k)} \in (\mathbb{R}^N)^{\otimes k}$ and $\boldsymbol{T} \in (\mathbb{R}^N)^{\otimes (k-1)}$, we denote by $\boldsymbol{A}^{(k)}\{\boldsymbol{T}\} \in \mathbb{R}^N$ the vector with components

$$\boldsymbol{A}^{(k)}\{\boldsymbol{T}\}_i = \frac{1}{(k-1)!} \sum_{1 \leq i_1,\dots,i_{k-1} \leq N} A^{(k)}_{i,i_1,\dots,i_{k-1}} T_{i_1 \dots i_{k-1}}. \tag{A.3}$$

For $\boldsymbol{u} \in \mathbb{R}^N$ we denote by $\boldsymbol{A}^{(k)}\{\boldsymbol{u}\} = \boldsymbol{A}^{(k)}\{\boldsymbol{u}^{\otimes (k-1)}\}$ the vector with entries

$$\boldsymbol{A}^{(k)}\{\boldsymbol{u}\}_i = \frac{1}{(k-1)!} \sum_{1 \leq i_1,\dots,i_{k-1} \leq N} A^{(k)}_{i,i_1,\dots,i_{k-1}} u_{i_1} \dots u_{i_{k-1}}. \tag{A.4}$$

Note that for $\boldsymbol{A}^{(k)}$ as in (A.1), one has

$$\boldsymbol{A}^{(k)}\{\boldsymbol{u}\} = \nabla H_{N,k}(\boldsymbol{u})$$

where $H_{N,k}$ denotes the part of H_N of total degree k.

For $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^N$ we recall from Sect. 1.3 the notations

$$\begin{aligned} \langle \boldsymbol{v} \rangle_N &= N^{-1} \sum_{i \leq N} v_i, \\ \langle \boldsymbol{u}, \boldsymbol{v} \rangle_N &= N^{-1} \sum_{i \leq N} u_i v_i = \langle \vec{\lambda}, \vec{R}(\boldsymbol{u}, \boldsymbol{v}) \rangle, \\ \|\boldsymbol{u}\|_N &= \langle \boldsymbol{u}, \boldsymbol{u} \rangle_N^{1/2} = \sqrt{\sum_s \lambda_s R_s(\boldsymbol{u}, \boldsymbol{u})}. \end{aligned}$$

Given functions $f_{t,s} : \mathbb{R}^{t+1} \to \mathbb{R}$ of $t+1$ variables for each $s \in \mathcal{S}$, and $\boldsymbol{v}^0, \boldsymbol{v}^1, \ldots, \boldsymbol{v}^t \in \mathbb{R}^N$, we define $f_t(\boldsymbol{v}^0, \boldsymbol{v}^1, \ldots, \boldsymbol{v}^t) \in \mathbb{R}^N$ component-wise via

$$f_t(\boldsymbol{v}^0, \boldsymbol{v}^1, \ldots, \boldsymbol{v}^t)_i = f_{t,s(i)}(v_i^0, \ldots, v_i^t). \tag{A.5}$$

Finally, for a sequence of vectors $\boldsymbol{w}^0, \boldsymbol{w}^1, \ldots$, we write $\boldsymbol{w}^{\le t} = (\boldsymbol{w}^0, \boldsymbol{w}^1, \ldots, \boldsymbol{w}^t)$.

To deduce the state evolution result for mixed tensors, we analyze a slightly more general iteration where each homogenous k-tensor is tracked separately, while restricting ourselves to the case where the mixture ξ has finitely many components: $\gamma_{s_1,\ldots,s_k} = 0$ for all $(s_1, \ldots, s_k) \in \mathcal{S}^k$ for all $k \ge D+1$ for some fixed $D \ge 2$. We then proceed by an approximation argument to extend the convergence to the general case $D = \infty$.

We begin by introducing the Gaussian process that captures the asymptotic behavior of AMP. Define ξ^k to be the degree k part of ξ, and

$$\xi^{k,s} = \frac{1}{\lambda_s} \partial_{x_s} \xi^k(x_1, \ldots, x_r)$$

the degree $k-1$ part of ξ^s.

An AMP iteration is specified by Lipschitz functions $f_{t,s} : \mathbb{R}^{2(t+1)} \to \mathbb{R}$ for each $(t, s) \in \mathbb{N} \times \mathcal{S}$.[3] For each iteration t, the state of the algorithm is given by vectors $\boldsymbol{w}^t \in \mathbb{R}^N$, and $\boldsymbol{z}^{k,t} \in \mathbb{R}^N$, with $k \in \{2, \ldots, D\}$. Moreover for each t, there is also an external randomness vector $\boldsymbol{e}^t \in \mathbb{R}^N$ with independent coordinates $e_i^t \sim \mu_{t,s(i)}$ from deterministic probability distributions $\left(\mu_{t,s}\right)_{t\ge 0, s\in\mathcal{S}}$ with finite second moment. We now start to define the AMP iteration steps (the definition finishes at (A.11)). A single step is given by

$$\mathsf{AMP}_t\left(\boldsymbol{w}^0, \ldots, \boldsymbol{w}^t; \boldsymbol{e}^0, \ldots, \boldsymbol{e}^t\right)_k \equiv \boldsymbol{A}^{(k)}\{\boldsymbol{f}_t\} - \sum_{t' \le t} d_{t,t',k} \diamond \boldsymbol{f}_{t'-1}, \tag{A.6}$$

$$\boldsymbol{f}_t = f_t(\boldsymbol{w}^0, \ldots, \boldsymbol{w}^t; \boldsymbol{e}^0, \ldots, \boldsymbol{e}^t), \tag{A.7}$$

$$d_{t,t',k,s} \equiv \left(\sum_{s' \in \mathcal{S}} \partial_{x_{s'}} \xi^{k,s} \left(\left(\mathbb{E}\left[F_{t,s} F_{t'-1,s} \right] \right)_{s\in\mathcal{S}} \right) \times \mathbb{E}\left[\partial_{W_{s'}^{t'}} F_{t,s'} \right] \right), \tag{A.8}$$

$$F_{t,s'} \equiv f_{t,s'}\left(W_{s'}^0, W_{s'}^1, \ldots, W_{s'}^t; E_{s'}^0, \ldots, E_{s'}^t\right). \tag{A.9}$$

A general multi-species tensor AMP algorithm then takes the form:

$$\boldsymbol{w}^t = \sum_{2 \le k \le D} \boldsymbol{z}^{k,t}, \qquad \boldsymbol{z}^{k,t+1} = \mathsf{AMP}_t\left(\boldsymbol{w}^0, \ldots, \boldsymbol{w}^t; \boldsymbol{e}^0, \ldots, \boldsymbol{e}^t\right)_k. \tag{A.10}$$

For the right-hand side of (A.9) to make sense, we must define for each $t \ge 0$ and $s \in \mathcal{S}$ a distribution over sequences $(W_s^0, \ldots, W_s^t; E_s^0, \ldots, E_s^t)$. The latter variables $E_s^{t'} \sim \mu_{t',s}$ are simply taken independent of each other and all other variables. The construction of the W variables is recursive across t as follows. For each $2 \le k \le D$ and $s \in \mathcal{S}$, we let $U_s^{k,0} \sim \nu_{k,s}$ and construct a centered Gaussian process

$$(U_s^{k,1}, U_s^{k,1}, \ldots, U_s^{k,t})$$

[3] The unusual factor 2 in the exponent comes from the external randomness vectors $\boldsymbol{e}^1, \ldots, \boldsymbol{e}^t$.

which is independent of $U_s^{k,0}$. The variables $U_s^{k,t}$ and $U_{s'}^{k',t'}$ are independent unless $(k,s) = (k',s')$. It remains to specify the covariance of $(U_s^{k,t})_{1\le t\le T}$ which is given recursively by:

$$\begin{aligned}\mathbb{E}\left[U_s^{k,t+1}U_s^{k,t'+1}\right] &= \xi^{k,s}(\Sigma^{k,t,t'});\\ \Sigma_s^{k,t,t'} &= \mathbb{E}\left[f_{t,s}(W_s^0,\dots,W_s^t;E_s^0,\dots,E_s^t)f_{t',s}(W_s^0,\dots,W_s^{t'};E_s^0,\dots,E_s^{t'})\right],\\ &\forall s\in\mathcal{S}\\ W_s^t &\equiv \sum_{2\le k\le D} U_s^{k,t}. \end{aligned} \tag{A.11}$$

The main result, an extension of Proposition 2.2, follows. Below we use $\mathbb{W}_2$ to denote the Wasserstein-2 distance between probability measures on Euclidean space in any dimension. We say a function $\psi:\mathbb{R}^d\to\mathbb{R}$ is **pseudo-Lipschitz** if

$$|\psi(\boldsymbol{w})-\psi(\boldsymbol{y})|\le C(1+\|\boldsymbol{w}\|+\|\boldsymbol{y}\|)\cdot\|\boldsymbol{w}-\boldsymbol{y}\|,\quad \forall \boldsymbol{w},\boldsymbol{y}\in\mathbb{R}^d.$$

Theorem 2 (State Evolution for AMP) *Let $\{\boldsymbol{G}^{(k)}\}_{k\ge 2}$ be independent standard Gaussian tensors with $\boldsymbol{G}^{(k)}\in(\mathbb{R}^N)^{\otimes k}$, and define $\boldsymbol{A}^{(k)}$ as in (A.2). Fix a sequence of Lipschitz functions $f_{t,s}:\mathbb{R}^{k+1}\to\mathbb{R}$. Let $\boldsymbol{z}^{2,0},\dots\boldsymbol{z}^{D,0}\in\mathbb{R}^N$ be deterministic vectors and $\boldsymbol{w}^0=\sum_{2\le k\le D}\boldsymbol{z}^{k,0}$. Assume that for each $s\in\mathcal{S}$, the empirical distribution of the vectors*

$$(z_i^{2,0},\dots z_i^{D,0}),\quad i\in\mathcal{I}_s$$

converges in $\mathbb{W}_2(\mathbb{R}^{D-1})$ distance to the law of the vector $(U_s^{k,0})_{2\le k\le D}$.

Let $\boldsymbol{w}^t$, $\boldsymbol{z}^{k,t}$, $t\ge 1$ be given by the tensor AMP *iteration. Then, for all $s\in\mathcal{S}$ and $T\ge 1$ and for any pseudo-Lipschitz functions $\psi:\mathbb{R}^{D\times T}\to\mathbb{R}$ and $\widetilde{\psi}:\mathbb{R}^T\to\mathbb{R}$, we have*

$$\operatorname*{p-lim}_{N\to\infty}\frac{1}{N_s}\sum_{i\in\mathcal{I}_s}\psi\left((z_i^{k,t})_{k\le D,t\le T}\right)=\mathbb{E}\left[\psi\left((U_s^{k,t})_{2\le k\le D,t\le T}\right)\right]; \tag{A.12}$$

$$\operatorname*{p-lim}_{N\to\infty}\frac{1}{N_s}\sum_{i\in\mathcal{I}_s}\widetilde{\psi}\left((w_i^t)_{t\le T}\right)=\mathbb{E}\left[\widetilde{\psi}\left((W_s^t)_{t\le T}\right)\right]. \tag{A.13}$$

Note that (A.13) (which concerns the actual AMP iterates $\boldsymbol{w}^t$) is a special case of (A.12) (which is more convenient to prove). Indeed one can take $\psi\left((z^{k,t})_{k\le D,t\le T}\right)=\widetilde{\psi}\left(\left(\sum_{k\le D}z^{k,t}\right)_{t\le T}\right)$. In the special case that $c_k=0$ for all $k\ge D+1$, Proposition 2.2 follows immediately from Theorem 2 by baking the contribution of $\boldsymbol{h}$ explicitly into f_t (since we require $k\ge 2$ above). Proposition 2.2 for non-polynomial ξ follows by a standard approximation argument outlined at the end of Sect. 1. For the remainder of this appendix we thus focus on establishing (A.12).

A.1: Further Definitions

We define the notations

$$\begin{aligned}\boldsymbol{W}_t&=\left[\boldsymbol{w}^0\mid\boldsymbol{w}^1\mid\dots\mid\boldsymbol{w}^t\right],\\ \boldsymbol{E}_t&=\left[\boldsymbol{e}^0\mid\boldsymbol{e}^1\mid\dots\mid\boldsymbol{e}^t\right],\\ \boldsymbol{Z}_{k,t}&=\left[\boldsymbol{z}^{k,0}\mid\boldsymbol{z}^{k,1}\mid\dots\mid\boldsymbol{z}^{k,t}\right].\end{aligned}$$

Given a $N \times (t+1)$ matrix such as $\boldsymbol{W}_t$, and a tensor $\boldsymbol{A}^{(k)} \in (\mathbb{R}^N)^{\otimes k}$, we write $\boldsymbol{A}^{(k)}\{\boldsymbol{W}_t\}$ for the $N \times (t+1)$ matrix with columns $\boldsymbol{A}^{(k)}\{\boldsymbol{w}^0\}$, …, $\boldsymbol{A}^{(k)}\{\boldsymbol{w}^t\}$:

$$\boldsymbol{A}^{(k)}\{\boldsymbol{W}_t\} = \Big[\boldsymbol{A}^{(k)}\{\boldsymbol{w}^0\}\Big|\boldsymbol{A}^{(k)}\{\boldsymbol{w}^1\}\Big|\dots\Big|\boldsymbol{A}^{(k)}\{\boldsymbol{w}^t\}\Big].$$

We will write $\boldsymbol{f}_t = f_t(\boldsymbol{W}_t, \boldsymbol{E}_t) = f_t(\boldsymbol{w}^0, \dots, \boldsymbol{w}^t, \boldsymbol{e}^0, \dots, \boldsymbol{e}^t)$ and also set

$$\boldsymbol{y}_{k,t+1}(\boldsymbol{Z}_{k,t}) = \boldsymbol{A}_k\left\{\boldsymbol{f}_t\right\} = \boldsymbol{z}^{k,t+1} + \sum_{t_1 \le t} d_{t,t_1,k} \diamond \boldsymbol{f}_{t_1-1}, \tag{A.14}$$

$$\boldsymbol{Y}_{k,t} = [\boldsymbol{y}_{k,1}|\dots|\boldsymbol{y}_{k,t}],$$

$$\boldsymbol{y}_t(\boldsymbol{Z}_{k,t}) = \sum_{2 \le k \le D} \boldsymbol{y}_{k,t}(\boldsymbol{Z}_{k,t}). \tag{A.15}$$

We also define an associated $(t+1) \times (t+1)$ Gram matrix $\boldsymbol{G}_{\xi^{k,s}} = \boldsymbol{G}_{\xi^{k,s},t}$ via

$$(\boldsymbol{G}_{\xi^{k,s}})_{t_1,t_2} = \xi^{k,s}\big(\vec{R}(\boldsymbol{f}_{t_1}, \boldsymbol{f}_{t_2})\big). \tag{A.16}$$

The dependence of $\boldsymbol{G}_{\xi^{k,s},t}$ on t will often be suppressed (this dependence is relevant when inverting the matrix $\boldsymbol{G}_{\xi^{k,s},t}$ but not for defining individual entries). Finally, we let $\mathcal{F}_t$ denote the σ-algebra generated by all iterates up to time t:

$$\mathcal{F}_t = \sigma\big(\{\boldsymbol{z}_{k,t_1}, \boldsymbol{w}^{t_1}, \boldsymbol{e}^{t_1}, \boldsymbol{f}_{t_1}\}_{k \le D, t_1 \le t}\big). \tag{A.17}$$

Throughout the proof of state evolution we make the following simplifying assumptions:

Assumption 2 ξ is a degree D polynomial with all coefficients $\gamma_{s_1,\dots,s_k}$ for $2 \le k \le D$ strictly positive.

Assumption 3 Each matrix $\boldsymbol{G}_{\xi^{k,s},t}$ is well-conditioned, i.e.

$$C^{-1} \le \sigma_{\min}(\boldsymbol{G}_{\xi^{k,s},t}) \le \sigma_{\max}(\boldsymbol{G}_{\xi^{k,s},t}) \le C$$

for all $t \le T$. Here $\boldsymbol{G}_{\xi^{k,s},t}$ is defined based on iterates that will appear in Theorems 3 and A.6. The same holds for $\mathcal{L}_{k,t}$ as defined in (A.34).

It is a standard argument that to establish Proposition 2.2, it suffices to do so under the above assumptions. The reason is that one can always slightly perturb both ξ and the nonlinearities $f_{t,s}$ to ensure the assumptions hold. Then suitable continuity properties suffice to transfer all asymptotic guarantees. We refer the reader to [4, Appendices A.8 and A.9] for the arguments in the single-species case, still in the generality of mixed tensors. (In the more common setting $D = 2$ of just a random matrix this step is also common for state evolution proofs, see e.g. [18, Sect. 4.2.1].) The corresponding extension in our setting is completely analogous and omitted.

A.2: Preliminary Lemmas

The next lemma has several parts. All are elementary Gaussian calculations so their proofs are omitted.

Lemma A.1 *For any deterministic $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^N$ and $\boldsymbol{A}^{(k)}$ defined by (A.2) we have:*

1. *Letting* $g_0 \sim \mathsf{N}(0,1)$ *be independent of* $\boldsymbol{g} \sim \mathsf{N}(0, \boldsymbol{I}_N)$*, we have*

$$\boldsymbol{A}^{(k)}\{\boldsymbol{u}\} \stackrel{\mathrm{d}}{=} \sum_{s\in\mathcal{S}} \boldsymbol{g}_s \sqrt{\xi^{k,s}(\vec{R}(\boldsymbol{u},\boldsymbol{u}))} + \frac{g_0}{\sqrt{N}} \sum_{s\in\mathcal{S}} \boldsymbol{u}_s \sqrt{\sum_{s'\in\mathcal{S}} \partial_{x_{s'}} \xi^{k,s}\Big(\vec{R}(\boldsymbol{u},\boldsymbol{u})\Big)}. \tag{A.18}$$

2. *Let* $g_0, g_1, \ldots, g_r \sim \mathsf{N}(0,1)$ *be independent. We have (jointly across* $s \in \mathcal{S}$*)*

$$\begin{aligned}\sqrt{\lambda_s N} R_s(\boldsymbol{v}, \boldsymbol{A}^{(k)}\{\boldsymbol{u}\}) \stackrel{\mathrm{d}}{=} &\sqrt{\xi^{k,s}(\vec{R}(\boldsymbol{u},\boldsymbol{u}))\cdot\vec{R}(\boldsymbol{v},\boldsymbol{v})}\, g_s \\ &+ \sqrt{\sum_{s'\in\mathcal{S}} \partial_{x_{s'}} \xi^{k,s}\Big(\vec{R}(\boldsymbol{u},\boldsymbol{u})\Big)}\, R_s(\boldsymbol{u},\boldsymbol{v}) \quad g_0. \end{aligned} \tag{A.19}$$

3. *For* $s \in \mathcal{S}$*:*

$$R\left(\boldsymbol{A}^{(k)}\{\boldsymbol{u}\}, \boldsymbol{A}^{(k)}\{\boldsymbol{v}\}\right)_s \simeq \xi^{k,s}\left(\vec{R}(\boldsymbol{u},\boldsymbol{v})\right).$$

4. *For a deterministic symmetric tensor* $\boldsymbol{T} \in (\mathbb{R}^N)^{\otimes k-1}$*, the vector* $\boldsymbol{A}^{(k)}\{\boldsymbol{T}\}$ *is centered Gaussian. Its covariance is given by*

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{A}^{(k)}\{\boldsymbol{T}\}_i \boldsymbol{A}^{(k)}\{\boldsymbol{T}\}_j\right] = &\langle \xi^{k,s(i)} \diamond \boldsymbol{T},\ \boldsymbol{T}\rangle_N \cdot 1\{i=j\} \\ &+ \frac{k(k-1)}{N^{k-1}} \sum_{i_1,\ldots,i_{k-2}=1}^{N} \gamma_{i,i_1,\ldots,i_{k-2}} \gamma_{j,i_1,\ldots,i_{k-2}} T_{i,i_1,\ldots,i_{k-2}} T_{j,i_1,\ldots,i_{k-2}}.\end{aligned}$$

5. *Let* $\boldsymbol{P} \in \mathbb{R}^{N\times N}$ *be the orthogonal projection onto a (deterministic) subspace* $S \subseteq \mathbb{R}^N$ *with* $d = \dim(S) = O(1)$*. Then*

$$\|\boldsymbol{P}\boldsymbol{G}^{(k)}\{\boldsymbol{u}\} - \boldsymbol{G}^{(k)}\{\boldsymbol{u}\}\|_2 / \|\boldsymbol{G}^{(k)}\{\boldsymbol{u}\}\|_2 \simeq 0.$$

We next develop a formula for the conditional expectation of a Gaussian tensor $\boldsymbol{A}^{(k)}$ given a collection of linear observations. We set $\boldsymbol{D}$ to be the $t\times t\times t$ tensor with entries $D_{ijk} = 1$ if $i = j = k$ and $D_{ijk} = 0$ otherwise.

Lemma A.2 *Recalling* (A.17)*, let* $\mathbb{E}\{\boldsymbol{A}^{(k)}|\mathcal{F}_t\}$*. Equivalently* $\mathbb{E}\{\boldsymbol{A}^{(k)}|\mathcal{F}_t\}$ *is the conditional expectation of* $\boldsymbol{A}^{(k)}$ *given the linear-in-*$\boldsymbol{A}^{(k)}$ *observations*

$$\boldsymbol{A}^{(k)}\{\boldsymbol{f}_{t'}\} = \boldsymbol{y}_{k,t'+1} \quad \text{for } s \in \{0,\ldots,t-1\}. \tag{A.20}$$

Then we have for $i_1, i_2, \ldots, i_k \le n$*,*

$$\begin{aligned}&\mathbb{E}[\boldsymbol{A}^{(k)}|\mathcal{F}_t]_{i_1,i_2,\ldots,i_k} \\ &= \sum_{j=1}^{k} \sum_{0\le t_1,t_2\le t-1} (\widehat{\boldsymbol{Z}}_{k,t})_{i_j,t_2} \cdot (\boldsymbol{G}^{-1}_{\xi^{k,s},t-1})_{t_2,t_1} \cdot (\boldsymbol{f}_{t_1,i_1}\cdots\boldsymbol{f}_{t_1,i_{j-1}}\boldsymbol{f}_{t_1,i_{j+1}}\cdots\boldsymbol{f}_{t_1,i_k}).\end{aligned} \tag{A.21}$$

Here, the matrix $\widehat{\boldsymbol{Z}}_{k,t} \in \mathbb{R}^{N\times t}$ *is defined as the solution of a system of linear equations as follows. Define the linear operator* $\mathcal{T}_{k,t} : \mathbb{R}^{N\times t} \to \mathbb{R}^{N\times t}$ *by letting, for* $i \le N$*,* $0 \le t_3 \le t-1$*:*

$$\begin{aligned}[\mathcal{T}_{k,t}(\boldsymbol{Z})]_{i,t_3} = &\sum_{j=1}^{N} \sum_{0\le t_1,t_2\le t-1} (\boldsymbol{f}_{t_2})_i (\boldsymbol{f}_{t_2})_j \\ &\times \left((\boldsymbol{G}^{-1}_{\xi^{k,s(i)},t-1})_{t_2,t_1} \partial_{s(j)} \xi^{k,s(i)}\big(\vec{R}(\boldsymbol{f}_{t_2},\boldsymbol{f}_{t_3})\big)\right) \diamond (\boldsymbol{Z})_{j,t_1}.\end{aligned} \tag{A.22}$$

Then $\widehat{\boldsymbol{Z}}_{k,t}$ is the unique solution of the following linear equation (with $\boldsymbol{Y}_{k,t}$ defined as per (A.14))

$$\widehat{\boldsymbol{Z}}_{k,t} + \mathcal{T}_{k,t}(\widehat{\boldsymbol{Z}}_{k,t}) = \boldsymbol{Y}_{k,t}. \tag{A.23}$$

(Here, $\widehat{\boldsymbol{Z}}_{k,t} = [\hat{z}_{k,0}, \dots, \hat{z}_{k,t-1}]$ and $\boldsymbol{Y}_{k,t} = [\hat{\boldsymbol{y}}_{k,1}, \dots, \hat{\boldsymbol{y}}_{k,t}]$ have dimensions $N \times t$.)

The above formulas for $\mathbb{E}[\boldsymbol{A}^{(k)}|\mathcal{F}_t]$ and $\mathcal{T}_{k,t}$ are rather complicated. In [4, Appendix A] the reader may find helpful tensor network diagrams for the single-species case. Unfortunately it is less clear how to draw a corresponding tensor network with multiple species.

Proof of Lemma A.2 Let $\mathcal{V}_{k,t}$ be the affine space of symmetric tensors satisfying the constraint (A.20). The conditional expectation $\mathbb{E}[\boldsymbol{A}^{(k)}|\mathcal{F}_t]$ is the tensor with minimum weighted Frobenius norm $\|\cdot\|_{F,\xi^k}$ in the affine space $\mathcal{V}_{k,t}$, given by

$$\|\boldsymbol{A}\|_{F,\xi^k}^2 = \langle (\Gamma^{(k)})^{-1} \diamond \boldsymbol{A}, (\Gamma^{(k)})^{-1} \diamond \boldsymbol{A} \rangle. \tag{A.24}$$

Here $(\Gamma^{(k)})^{-1}$ is the entry-wise inverse of $\Gamma^{(k)}$, which exists by Assumption 2.

By Lagrange multipliers, there exist vectors $\boldsymbol{m}^1, \dots, \boldsymbol{m}^t \in \mathbb{R}^N$ such that $\mathbb{E}[\boldsymbol{A}^{(k)}|\mathcal{F}_t] = \widehat{\boldsymbol{A}}^{(k)}$ equals

$$\widehat{\boldsymbol{A}}_t^{(k)} \equiv \Gamma^{(k)} \diamond \sum_{t'=0}^{t-1} \sum_{j=1}^{k} \underbrace{\boldsymbol{f}_{t'} \otimes \cdots \otimes \boldsymbol{f}_{t'}}_{j-1 \text{ times}} \otimes \boldsymbol{m}^{t'} \otimes \underbrace{\boldsymbol{f}_{t'} \otimes \cdots \otimes \boldsymbol{f}_{t'}}_{k-j \text{ times}}. \tag{A.25}$$

Also by Lagrange multipliers, if a tensor $\widehat{\boldsymbol{A}}^{(k)}$ is of this form (for some choice of vectors $\boldsymbol{m}^1, \dots, \boldsymbol{m}^t$) and satisfies the constraints $\widehat{\boldsymbol{A}}^{(k)}\{\boldsymbol{f}_{t'}\} = \boldsymbol{y}_{k,t'+1}$ for $s < t$, then this tensor is unique and equals $\mathbb{E}[\boldsymbol{A}^{(k)}|\mathcal{F}_t]$. Without loss of generality, we write

$$\boldsymbol{m}_i^{t_1} = \sum_{t_2=0}^{t-1} (\boldsymbol{G}^{-1}_{\xi^{k,s(i)},t-1})_{t_1,t_2} \hat{z}_i^{t_2}, \quad \widehat{\boldsymbol{Z}}_{k,t} = [\hat{z}^1 | \dots | \hat{z}^t]. \tag{A.26}$$

By direct calculation we obtain that for each $i \in [N]$,

$$\begin{aligned}
\big(\widehat{\boldsymbol{A}}_t^{(k)}\{\boldsymbol{f}_{t_1}\}\big)_i &= \sum_{t_2=0}^{t-1} (\boldsymbol{G}_{\xi^{k,s(i)},t-1})_{t_1,t_2} (\boldsymbol{m}^{t_2})_i \\
&\quad + \sum_{t_2=0}^{t-1} \sum_{s' \in \mathcal{S}} \Big(\partial_{s'} \xi^{k,s(i)}(\vec{R}(\boldsymbol{f}_{t_1}, \boldsymbol{f}_{t_2})) R_{s'}(\boldsymbol{f}_{t_1}, \boldsymbol{m}^{t_2}) \Big) (\boldsymbol{f}_{t_2})_i \\
f &= (\hat{z}_{t_1})_i + \sum_{t_2=0}^{t-1} \sum_{s' \in \mathcal{S}} \Big(\partial_{s'} \xi^{k,s(i)}(\vec{R}(\boldsymbol{f}_{t_1}, \boldsymbol{f}_{t_2})) R_{s'}(\boldsymbol{f}_{t_1}, \boldsymbol{m}^{t_2}) \Big) (\boldsymbol{f}_{t_2})_i .
\end{aligned} \tag{A.27}$$

We next stack these vectors as columns of an $N \times t$ matrix. The first term yields $\widehat{\boldsymbol{Z}}_{k,t}$. Moreover the second term coincides with $\mathcal{T}_{k,t}(\widehat{\boldsymbol{Z}}_{k,t})$ by rearranging the order of sums in (A.27). Hence

$$\big[\widehat{\boldsymbol{A}}_t^{(k)}\{\boldsymbol{f}_0\}, \dots, \widehat{\boldsymbol{A}}_t^{(k)}\{\boldsymbol{f}_{t-1}\}\big] = \widehat{\boldsymbol{Z}}_{k,t} + \mathcal{T}_{k,t}(\widehat{\boldsymbol{Z}}_{k,t}). \tag{A.28}$$

This in turn implies that the equation determining $\widehat{\boldsymbol{Z}}_{k,t}$ takes the form (A.23). □

A.3: Long AMP

As an intermediate step towards proving Theorem 2, we introduce a new iteration that we call Long AMP (LAMP), following [8]. This iteration is less compact but simpler to analyze. For each $k \le D$, let $\mathcal{S}_{k,t} \subseteq (\mathbb{R}^N)^{\otimes k}$ be the linear subspace of tensors $\boldsymbol{T}$ that are symmetric and such that $\boldsymbol{T}\{\boldsymbol{f}_{t_1}\} = 0$ for all $t_1 < t$. We denote by $\mathcal{P}_t^{\perp}(\boldsymbol{A}^{(k)})$ the projection of $\boldsymbol{A}^{(k)}$ onto $\mathcal{S}_{k,t}$, in the inner product space (A.24) corresponding to $\Gamma^{(k)}$. We then define the LAMP mapping

$$\mathsf{LAMP}_t\left(\boldsymbol{v}^{\le t}\right)_k \equiv \mathcal{P}_t^{\perp}(\boldsymbol{A}^{(k)})\{\boldsymbol{f}_t\} + \sum_{0\le t_1\le t} h_{t,t_1-1,k} \diamond \boldsymbol{q}^{k,t_1}, \tag{A.29}$$

$$h_{t,t_1,k,s} \equiv \sum_{0\le t_2\le t-1} \left[\boldsymbol{G}^{-1}_{\xi^{k,s},t-1}\right]_{t_1,t_2}\left[\boldsymbol{G}_{\xi^{k,s},t}\right]_{t_2,t}, \quad h_{t,-1,k} = 0. \tag{A.30}$$

Here we use similar notations $\boldsymbol{f}_t = f_t(\boldsymbol{V}_t; \boldsymbol{E}_t)$ and $\boldsymbol{G}_{\xi^{k,s},t}$ as before (recall (A.16)), and take the vectors $\boldsymbol{e}^t$ as before. However the quantities $\boldsymbol{f}_t, \boldsymbol{G}_{\xi^{k,s},t}$ are now different: they are computed using the vectors $\boldsymbol{v}^0, \dots, \boldsymbol{v}^t$ using the recursion:

$$\boldsymbol{v}^t = \sum_{2\le k\le D} \boldsymbol{q}^{k,t}, \quad \boldsymbol{q}^{k,t+1} = \mathsf{LAMP}_t\left(\boldsymbol{v}^{\le t}\right)_k. \tag{A.31}$$

Following [4, 8], we first establish state evolution for LAMP (under the non-degeneracy Assumption 2), and then deduce the result for the original AMP. In analyzing LAMP we use notations analogous to the ones introduced for AMP. In particular:

$$\boldsymbol{V}_t = [\boldsymbol{v}_1|\boldsymbol{v}_2|\dots|\boldsymbol{v}_t] \tag{A.32}$$

$$\boldsymbol{Q}_{k,t} = [\boldsymbol{q}^{\otimes k}_{k,1}|\boldsymbol{q}^{\otimes k}_{k,2}|\dots|\boldsymbol{q}^{\otimes k}_{k,t}]. \tag{A.33}$$

A.4: State Evolution for Long AMP

Theorem 3 *Under the assumptions of Theorem 2, let $\boldsymbol{q}^{2,0}, \dots \boldsymbol{q}^{D,0} \in \mathbb{R}^N$ be deterministic vectors and $\boldsymbol{v}^0 = \sum_{2\le k\le D} \boldsymbol{q}^{k,0}$. Assume that the uniform empirical distribution of the N vectors $\{(q_i^{2,0}, \dots, q_i^{D,0})\}_{i\le N}$ converges in $\mathbb{W}_2$ distance to the law of the vector $(U^{k,0})_{2\le k\le D}$.*

Further we assume there is a constant $C < \infty$ such that for all $t \le T$:

(i) *The matrices $\boldsymbol{G}_{\xi^{k,s},t} = \boldsymbol{G}_{\xi^{k,s},t}(\boldsymbol{V})$ are uniformly well-conditioned as guaranteed by Assumption 3.*

(ii) *Let the linear operator $\mathcal{T}_{k,t} : \mathbb{R}^{N\times t} \to \mathbb{R}^{N\times t}$ be defined as per (A.22), with $\boldsymbol{G}_{\xi^{k,s},t} = \boldsymbol{G}_{\xi^{k,s},t}(\boldsymbol{V}, \boldsymbol{E})$, and $\boldsymbol{f}_t = f_t(\boldsymbol{V}, \boldsymbol{E})$, and define*

$$\mathcal{L}_{k,t} = \mathbf{1} + \mathcal{T}_{k,t}. \tag{A.34}$$

Then $C^{-1} \le \sigma_{\min}(\mathcal{L}_{k,t}) \le \sigma_{\max}(\mathcal{L}_{k,t}) \le C$.

Then the following statements hold for any $t \le T$ and sufficiently large N:

(a) *Correct conditional law:*

$$\boldsymbol{q}^{k,t+1}|_{\mathcal{F}_t} \stackrel{\mathrm{d}}{=} \mathbb{E}[\boldsymbol{q}^{k,t+1}|\mathcal{F}_t] + \mathcal{P}_t^{\perp}(\widetilde{\boldsymbol{A}}^{(k)})\{\boldsymbol{f}_t\}. \tag{A.35}$$

where $\widetilde{\boldsymbol{A}}^{(k)}$ is a symmetric tensor distributed identically to $\boldsymbol{A}^{(k)}$ and independent of everything else, and $\mathcal{P}_t^{\perp}$ is the projection onto the subspace $\mathcal{S}_{k,t}$ defined in Sect. 1. Further

$$\mathbb{E}[\boldsymbol{q}^{k,t+1}|\mathcal{F}_t] = \sum_{s\in\mathcal{S}} \sum_{0\le t_1\le t} h_{t,t_1-1,k,s}\boldsymbol{q}_s^{k,t_1}. \tag{A.36}$$

Moreover, the vectors $(\boldsymbol{q}^{k,t+1})_{2\le k\le D}$ *are conditionally independent given* $\mathcal{F}_t$.

(b) *Approximate isometry: we have*

$$R_s(\boldsymbol{q}^{k,t_1+1}, \boldsymbol{q}^{k,t_2+1}) \simeq \xi^{k,s}\left(\vec{R}(\boldsymbol{f}_{t_1}, \boldsymbol{f}_{t_2})\right), \tag{A.37}$$

$$R_s(\boldsymbol{v}^{t_1+1}, \boldsymbol{v}^{t_2+1}) \simeq \xi^{s}\left(\vec{R}(\boldsymbol{f}_{t_1}, \boldsymbol{f}_{t_2})\right). \tag{A.38}$$

Moreover, both sides converge in probability to constants as $N \to \infty$, *and for* $k_1 \neq k_2$ *and any* (t_1, t_2) *and* $s \in \mathcal{S}$,

$$R_s(\boldsymbol{q}^{k_1,t_1}, \boldsymbol{q}^{k_2,t_2}) \simeq 0. \tag{A.39}$$

(c) *State evolution: for each* $s \in \mathcal{S}$ *and any pseudo-Lipschitz function* $\psi : \mathbb{R}^{D\times 2(t+1)} \to \mathbb{R}$, *we have*

$$\operatorname*{p-lim}_{N\to\infty} \frac{1}{N_s} \sum_{i\in\mathcal{I}_s} \psi\left((q_i^{k,t'})_{k\le D, t'\le t}; (e_i^t)_{t'\le t}\right) = \mathbb{E}\left\{\psi\left((U_s^{k,t'})_{2\le k\le D, t'\le t}; (E_s^{t'})_{t'\le t}\right)\right\}. \tag{A.40}$$

where $(U_s^{k,t})_{k\le D, 1\le t\le T}$ *is the centered Gaussian process defined in the statement of Theorem 2.*

In the next subsection, we will prove these statements by induction on t. The crucial point we exploit is the representation (a). We emphasize that the iteration number t is bounded as $N \to \infty$; therefore all numerical quantities not depending on N (but possibly on t) will be treated as constants.

A.5: Proof of Theorem 3

The proof will be by induction over t. The base case is clear, (e.g. see Proposition A.4) and we focus on the inductive step. We assume the statements above for $t-1$ and prove them for t.

A.5.1: Proof of (*a*)

Note that $\mathcal{P}_t^{\perp}(\boldsymbol{A}^{(k)})$ is by construction independent of $\mathcal{F}_t$, and therefore we can replace $\boldsymbol{A}^{(k)}$ by a fresh independent matrix in (A.29), whence (A.35) follows. The equality (A.36) holds by definition of the iteration.

A.5.2: Proof of (*b*): Approximate isometry

We will repeatedly apply Lemma A.1. We start with (A.37). As we are inducting on t, we may limit ourselves to considering overlaps $\vec{R}(\boldsymbol{q}^{k,t+1}, \boldsymbol{q}^{k,t_1+1})$, for $t_1 \le t$.

Define the tensor $\Gamma^{(k),\nabla} \in (\mathbb{R}^{\mathcal{S}}_{\ge 0})^{\otimes(k-1)}$ by

$$\Gamma^{(k),\nabla}_{s_1,\dots,s_{k-1}} = \sqrt{k \sum_{s\in\mathcal{S}} \lambda_s \left(\Gamma^{(k)}_{s,s_1,\dots,s_{k-1}}\right)^2}. \tag{A.41}$$

We choose

$$(\boldsymbol{f}_t^{\otimes k-1})_{\|} \in \operatorname{span}\left(\boldsymbol{f}_{t_1}^{\otimes k-1}\right)_{t_1<t}$$

such that

$$\Gamma^{(k),\nabla} \diamond \left(\boldsymbol{f}_t^{\otimes k-1}\right)_{\|}$$

is the orthogonal projection of $\Gamma^{(k),\nabla} \diamond \boldsymbol{f}_t^{\otimes k-1}$ onto

$$\text{span}\left(\Gamma^{(k),\nabla} \diamond \boldsymbol{f}_{t_1}^{\otimes k-1}\right)_{t_1<t}$$

and also set

$$(\boldsymbol{f}_t^{\otimes k-1})_{\perp} = \boldsymbol{f}_t^{\otimes k-1} - (\boldsymbol{f}_t^{\otimes k-1})_{\|}.$$

We will use (and soon after, prove) the following lemma.

Lemma A.3 *For all $t_1 \leq t_1$, we have*

$$\mathcal{P}_t^{\perp}(\widetilde{\boldsymbol{A}}^{(k)})\{(\boldsymbol{f}_t^{\otimes k-1})_{\perp}\} \simeq \widetilde{\boldsymbol{A}}^{(k)}\{(\boldsymbol{f}_t^{\otimes k-1})_{\perp}\}. \tag{A.42}$$

For $t_1 \leq t-1$, using Lemma A.1, point 2 implies

$$\vec{R}(\boldsymbol{q}^{k,t+1}, \boldsymbol{q}^{k,t_1+1}) \simeq \vec{R}\big(\mathbb{E}[\boldsymbol{q}^{k,t+1}|\mathcal{F}_t], \quad \boldsymbol{q}^{k,t_1+1}\big)$$

We next use the formula in (a) for $\mathbb{E}[\boldsymbol{q}^{k,t+1}|\mathcal{F}_t]$ together with the expression in (A.29). For each $s \in \mathcal{S}$:

$$\begin{aligned}
R\big(\mathbb{E}[\boldsymbol{q}^{k,t+1}|\mathcal{F}_t], \boldsymbol{q}^{k,t_1+1}\big)_s &\simeq R\Big(\sum_{0\leq t_2,t_3\leq t-1} \boldsymbol{q}^{k,t_3+1}(\boldsymbol{G}^{-1}_{\xi^{k,s},t-1})_{t_3,t_2} \quad \xi^{k,s}\big(\boldsymbol{f}_{t_2}, \boldsymbol{f}_t\big), \quad \boldsymbol{q}^{k,t_1+1}\Big)_s \\
&= \sum_{0\leq t_2,t_3\leq t-1} R\big(\boldsymbol{q}^{k,t_3+1}, \boldsymbol{q}^{k,t_1+1}\big)_s \; (\boldsymbol{G}^{-1}_{\xi^{k,s},t-1})_{t_3,t_2} \; \xi^{k,s}\big(\boldsymbol{f}_{t_2}, \boldsymbol{f}_t\big) \\
&\simeq \sum_{0\leq t_2,t_3\leq t-1} (\boldsymbol{G}_{\xi^{k,s},t-1})_{t_3,t_1} \; (\boldsymbol{G}^{-1}_{\xi^{k,s},t-1})_{t_3,t_2} \; \xi^{k,s}\big(\boldsymbol{f}_{t_2}, \boldsymbol{f}_t\big) \\
&= \big(\boldsymbol{G}_{\xi^{k,s},t-1} \times \boldsymbol{G}^{-1}_{\xi^{k,s},t-1} \times \boldsymbol{G}_{\xi^{k,s},t-1}\big)_{t_1,t} \\
&= (\boldsymbol{G}_{\xi^{k,s},t-1})_{t_1,t}.
\end{aligned} \tag{A.43}$$

Here (A.43) comes from the induction hypothesis (A.37) (and the symmetry of the matrix $\boldsymbol{G}_{\xi^{k,s},t-1}$ is used to obtain the next line). We next prove that (A.37) holds for $t_1 = t$. We have by definition of the projections that

$$\mathcal{P}_t^{\perp}(\widetilde{\boldsymbol{A}}^{(k)})\{\boldsymbol{f}_t\} = \mathcal{P}_t^{\perp}(\widetilde{\boldsymbol{A}}^{(k)})\{(\boldsymbol{f}_t^{\otimes k-1})_{\perp}\},$$

where the right-hand side is defined according to (A.3). Using (A.42) from Lemma A.3 as well as point 4 of Lemma A.1, we have

$$R\left(\mathcal{P}_t^{\perp}(\widetilde{\boldsymbol{A}}^{(k)})\{\boldsymbol{f}_t\}, \mathcal{P}_t^{\perp}(\widetilde{\boldsymbol{A}}^{(k)})\{\boldsymbol{f}_t\}\right)_s \simeq \xi^{(k,s)}\left(R\big((\boldsymbol{f}_t^{\otimes k-1})_{\perp}, (\boldsymbol{f}_t^{\otimes k-1})_{\perp}\big)\right). \tag{A.44}$$

Next, using (A.42) and Lemma A.1 (point 2), we obtain that for all $s \in \mathcal{S}$

$$R\big(\mathcal{P}_t^{\perp}(\widetilde{\boldsymbol{A}}^{(k)})\{\boldsymbol{f}_t\}, \mathbb{E}[\boldsymbol{q}^{k,t+1}|\mathcal{F}_t]\big)_s \simeq 0. \tag{A.45}$$

Moreover we recall that by the expression for $\mathbb{E}[\boldsymbol{q}^{k,t+1}|\mathcal{F}_t]$ from part (a),

$$R\left(\mathbb{E}[\boldsymbol{q}^{k,t+1}|\mathcal{F}_t], \mathbb{E}[\boldsymbol{q}^{k,t+1}|\mathcal{F}_t]\right) \simeq \xi^{k,s}\left(\vec{R}((\boldsymbol{f}_t^{\otimes k-1})_{\|}, (\boldsymbol{f}_t^{\otimes k-1})_{\|})\right). \tag{A.46}$$

The formula for linear regression implies

$$(\boldsymbol{f}_t^{\otimes k-1})_{\|} = \sum_{0\le t_1\le t-1} \alpha_{t_1,t}\Gamma^{(k,s)} \diamond \boldsymbol{f}_{t_1}^{\otimes k-1}, \tag{A.47}$$

$$\alpha_{t_1,t} = \sum_{0\le t_2\le t-1} (\boldsymbol{G}^{-1}_{\xi^{k,s},t-1})_{t_1,t_2}\langle \Gamma^{(k,s)} \diamond \boldsymbol{f}_{t_2}^{\otimes k-1}, \Gamma^{(k,s)} \diamond \boldsymbol{f}_{t}^{\otimes k-1}\rangle_N \tag{A.48}$$

$$= \sum_{0\le t_2\le t-1} (\boldsymbol{G}^{-1}_{\xi^{k,s},t-1})_{t_1,t_2}(\boldsymbol{G}_{\xi^{k,s},t})_{t_2,t}. \tag{A.49}$$

By part (b) of the inductive step, for $1 \le t_1, t_2 \le t-1$ we have

$$\xi^{k,s}\left(\vec{R}(\boldsymbol{f}_{t_2}, \boldsymbol{f}_{t_1})\right) \simeq R_s(\boldsymbol{q}_{k,t_2+1}, \boldsymbol{q}_{k,t_1+1}).$$

In particular the formulas (A.30) and (A.47) have asymptotically the same coefficients, and the overlap structure between the summands is identical. It follows that

$$R\left(\mathbb{E}[\boldsymbol{q}^{k,t+1}|\mathcal{F}_t], \mathbb{E}[\boldsymbol{q}^{k,t+1}|\mathcal{F}_t]\right) \simeq R\left((\boldsymbol{f}_t^{\otimes k-1})_{\|}, (\boldsymbol{f}_t^{\otimes k-1})_{\|}\right). \tag{A.50}$$

Using together Eqs. (A.44), (A.45), and (A.50), we get

$$\begin{aligned} R\left(\boldsymbol{q}^{k,t+1}, \boldsymbol{q}^{k,t+1}\right) &\simeq R\left(\mathbb{E}[\boldsymbol{q}^{k,t+1}|\mathcal{F}_t], \mathbb{E}[\boldsymbol{q}^{k,t+1}|\mathcal{F}_t]\right) + R\left((\boldsymbol{f}_t^{\otimes k-1})_{\perp}, (\boldsymbol{f}_t^{\otimes k-1})_{\perp}\right) \\ &\simeq R\left((\boldsymbol{f}_t^{\otimes k-1})_{\perp}, (\boldsymbol{f}_t^{\otimes k-1})_{\perp}\right) \\ &= \xi^{k,s}\left(\vec{R}(\boldsymbol{f}_t, \boldsymbol{f}_t)\right). \end{aligned}$$

This establishes (A.37).

Next consider (A.39), i.e., approximate orthogonality of $\boldsymbol{q}^{k,r}$ and $\boldsymbol{q}^{p',r}$ for $k \ne p'$. This follows easily from the representation in point (a) which, together with Lemma A.1, inductively implies that the iterates $\boldsymbol{q}^{s,k}$ for different k are approximately orthogonal. Finally, (A.38) follows directly from (A.37) and (A.39). We now prove Lemma A.3.

Proof of Lemma A.3 For convenience we write $\widetilde{\boldsymbol{A}} = \widetilde{\boldsymbol{A}}^{(k)}$. By Lagrange multipliers, there exist vectors $(\boldsymbol{\theta}_{t_1})_{t_1\le t-1}$ in $\mathbb{R}^N$ such that $\mathcal{P}_t^{\perp}(\widetilde{\boldsymbol{A}}) = \widetilde{\boldsymbol{A}} - \boldsymbol{Q}$, where

$$\boldsymbol{Q} = \left(\Gamma^{(k)}\right)^{\odot 2} \diamond \frac{(k-1)!}{N^{k-1}} \sum_{t_1=0}^{t-1}\sum_{j=1}^{k} \underbrace{\boldsymbol{f}_{t_1}\otimes\cdots\otimes\boldsymbol{f}_{t_1}}_{j-1 \text{ times}} \otimes \boldsymbol{\theta}_{t_1} \otimes \underbrace{\boldsymbol{f}_{t_1}\otimes\cdots\otimes\boldsymbol{f}_{t_1}}_{k-j \text{ times}}.$$

The vectors $(\boldsymbol{\theta}_{t_1})_{t_1\le t-1}$ are determined by the equations $\boldsymbol{Q}\{\boldsymbol{f}_{t_1}\} = \widetilde{\boldsymbol{A}}\{\boldsymbol{f}_{t_1}\}$ for all $t_1 \le t-1$. This expands (for each $t_1 \le t-1$) to

$$\sum_{t_2\le t-1} (\boldsymbol{G}_{\xi^{k,s},t-1})_{t_1,t_2} \diamond \boldsymbol{\theta}_{t_2} + \sum_{t_2\le t-1}\sum_{s'\in\mathcal{S}} \left(\partial_{s'}\xi^{k,s(i)}(\vec{R}(\boldsymbol{f}_{t_1}, \boldsymbol{f}_{t_2}))R_{s'}(\boldsymbol{f}_{t_1}, \boldsymbol{\theta}_{t_2})\right)\boldsymbol{f}_{t_2} = \widetilde{\boldsymbol{A}}\{\boldsymbol{f}_{t_1}\}.$$

Recall that we assume each $\boldsymbol{G}_{\xi^{k,s},t-1}$ is well-conditioned with high probability. Thus we can multiply the system of t equations above by $\boldsymbol{G}^{-1}_{\xi^{k,s},t-1}$ in the coordinates $\mathcal{I}_s$ for each $s \in \mathcal{S}$. For each $t_3 \le t-1$, we obtain:

$$\begin{aligned} &\boldsymbol{\theta}_{t_3} + \sum_{t_1,t_2<t}\left((\boldsymbol{G}^{-1}_{\xi^{k,s},t-1})_{t_1,t_3}\sum_{s'\in\mathcal{S}}\left(\partial_{s'}\xi^{k,s(i)}(\vec{R}(\boldsymbol{f}_{t_1}, \boldsymbol{f}_{t_2}))R_{s'}(\boldsymbol{f}_{t_1}, \boldsymbol{\theta}_{t_2})\right)\right)\boldsymbol{f}_{t_2} \\ &= \sum_{t_1<t}(\boldsymbol{G}^{-1}_{\xi^{k,s},t-1})_{t_1,t_3} \diamond \widetilde{\boldsymbol{A}}\{\boldsymbol{f}_{t_1}\}. \end{aligned} \tag{A.51}$$

Switching t_3 to t_1, we find

$$\begin{aligned}\boldsymbol{\theta}_{t_1} &= \boldsymbol{\theta}^0_{t_1} + \boldsymbol{\theta}^{\|}_{t_1},\\ \boldsymbol{\theta}^0_{t_1} &\equiv \sum_{t_2<t} (\boldsymbol{G}^{-1}_{\xi^{k,s},t-1})_{t_1,t_2} \diamond \widetilde{\boldsymbol{A}}\{\boldsymbol{f}_{t_2}\},\\ \boldsymbol{\theta}^{\|}_{t_1} &\in \operatorname{span}\left((\boldsymbol{f}_{t_2,s})_{t_2<t,s\in\mathcal{S}}\right). \end{aligned} \tag{A.52}$$

We claim that $\|\boldsymbol{\theta}^{\|}_{t_1}\|_N \simeq 0$, i.e., $\boldsymbol{\theta}_{t_1} \simeq \boldsymbol{\theta}^0_{t_1}$. Indeed, let $\boldsymbol{\Theta} \in \mathbb{R}^{N\times t}$ be the matrix with columns $(\boldsymbol{\theta}_{t_2})_{t_2<t}$, and $\boldsymbol{\Theta}^0$ the matrix with columns $(\boldsymbol{\theta}^0_{t_2})_{t_2<t}$. Then (A.51) can be written as

$$\mathcal{L}^{\mathsf{T}}_{k,t}(\boldsymbol{\Theta}) = \boldsymbol{\Theta}^0.$$

Here we recall $\mathcal{L}_{k,t} = \mathbf{1} + \mathcal{T}_{k,t}$ and $\mathcal{T}_{k,t} \in \mathbb{R}^{Nt\times Nt}$ is defined in (A.22). Substituting the decomposition $\boldsymbol{\Theta} = \boldsymbol{\Theta}^0 + \boldsymbol{\Theta}^{\|}$ in the above, we obtain

$$\mathcal{L}^{\mathsf{T}}_{k,t}(\boldsymbol{\Theta}^{\|}) = -\mathcal{T}^{\mathsf{T}}_{k,t}(\boldsymbol{\Theta}^0).$$

Recall that $\mathcal{L}_{k,t}$ is well-conditioned by Assumption 3. Therefore it remains to prove

$$\mathcal{T}^{\mathsf{T}}_{k,t}(\boldsymbol{\Theta}^0) \overset{?}{\simeq} 0. \tag{A.53}$$

Let $\boldsymbol{c}_0, \ldots, \boldsymbol{c}_{t-1} \in \mathbb{R}^N$ be the columns of $\mathcal{T}^{\mathsf{T}}_{k,t}(\boldsymbol{\Theta}^0)$. We first note that for all $t_1 \le t-1$ and $s \in \mathcal{S}$,

$$\boldsymbol{c}_{t_1,s} \in \operatorname{span}\big((\boldsymbol{f}_{t_2,s})_{t_2<t}\big).$$

Moreover the Gram matrix

$$\boldsymbol{G}_{1,t-1,s} = \big(R_s(\boldsymbol{f}_{t_1}, \boldsymbol{f}_{t_2})\big)_{t_1,t_2<t}$$

is well-conditioned for each $s \in \mathcal{S}$. Therefore it is sufficient to check that $R_s(\boldsymbol{f}_{t_1}, \boldsymbol{c}_{t_4}) \simeq 0$ for each $t_1, t_4 < t$ and $s \in \mathcal{S}$. Plugging in the definition (A.22), it remains to check that for $0 \le t_1, t_4 \le t-1$,

$$\sum_{t_2,t_3<t}\sum_{s''\in\mathcal{S}} \lambda_{s''}\vec{R}_s\left(\boldsymbol{f}_{t_1},\ (\boldsymbol{G}^{-1}_{\xi^{k,s''},t-1})_{t_2,t_3}\sum_{s'}\partial_{s'}\xi^{k,s''}(\vec{R}(\boldsymbol{f}_{t_4},\boldsymbol{f}_{t_3}))R_{s'}(\boldsymbol{f}_{t_4},\boldsymbol{\theta}^0_{t_2})\right) R_s(\boldsymbol{f}_{t_1},\boldsymbol{f}_{t_3}) \overset{?}{\simeq} 0.$$

Finally, this last claim follows by substituting the definition (A.52) of $\boldsymbol{\theta}^0_{t_2}$, and using the fact that

$$R_{s'}(\boldsymbol{f}_{t_4}, \widetilde{\boldsymbol{A}}\{\boldsymbol{f}_{t_2}\}) \simeq 0, \quad \forall\, t_4, t_2 \le t, \quad s' \in \mathcal{S}$$

which follows from Lemma A.1. Thus (A.53) is established.

We are now ready to prove Lemma A.3. First note that

$$\widetilde{\boldsymbol{A}}\{(\boldsymbol{f}_t^{\otimes k-1})_\perp\} - \mathcal{P}^\perp_t(\widetilde{\boldsymbol{A}})\{(\boldsymbol{f}_t^{\otimes k-1})_\perp\} = \boldsymbol{Q}\{(\boldsymbol{f}_t^{\otimes k-1})_\perp\} \tag{A.54}$$

decomposes into two types of terms based on the definition of $\boldsymbol{Q}$ above. Recalling (A.41), the first involves

$$\left\langle \Gamma^{(k),\nabla}\diamond(\boldsymbol{f}_{t_1}^{\otimes k-1})_\perp, \Gamma^{(k),\nabla}\diamond(\boldsymbol{f}_t^{\otimes k-1})_\perp\right\rangle_N \boldsymbol{\theta}_{t_1}$$

for $t_1 \le t-1$, which vanishes by the definition of $(\boldsymbol{f}_t^{\otimes k-1})_\perp$. The other terms take the form

$$\left\langle \Gamma^{(k),\nabla}\diamond(\boldsymbol{\theta}_{t_1}\otimes\boldsymbol{f}_{t_1}^{\otimes k-2}),\quad \Gamma^{(k),\nabla}\diamond(\boldsymbol{f}_t^{\otimes k-1})_\perp\right\rangle_N \boldsymbol{f}_{t_1}.$$

In particular, this means that to prove (A.54) vanishes, suffices to show

$$R\left(\boldsymbol{Q}\{(\boldsymbol{f}_t^{\otimes k-1})_\perp\}, \quad \boldsymbol{f}_{t_2}\right) = 0$$

for all $t_2 \le t$.

Note that by construction,

$$(\boldsymbol{f}_t^{\otimes k-1})_\perp = \sum_{t_1 \le t} b_{t_1} \boldsymbol{f}_{t_1}^{\otimes k-1}.$$

By the well-conditioning assumption, the b_{t_1} are bounded. Therefore it suffices to show that

$$R\left(\boldsymbol{Q}\{\boldsymbol{f}_{t_2}^{\otimes k-1}\}, \quad \boldsymbol{f}_{t_1}\right) \overset{?}{\simeq} 0, \quad \forall\, t_1 \le t-1, \quad t_2 \le t.$$

Finally note that each term in the left-hand side includes an overlap $R_s(\boldsymbol{\theta}_{t_1}, \boldsymbol{f}_{t_2})$. However these all vanish:

$$R_s(\boldsymbol{\theta}_{t_1}, \boldsymbol{f}_{t_2}) \simeq 0.$$

This is because we can substitute $\boldsymbol{\theta}_{t_1}$ with $\boldsymbol{\theta}^0_{t_1}$ as defined in (A.52) and use the fact that $\vec{R}(\widetilde{\boldsymbol{A}}\{\boldsymbol{f}_{t_3}\}, \boldsymbol{f}_t) \simeq 0$ which follows from Lemma A.1. This completes the proof. □

A.5.3: Proof of (*c*)

The base case of initialization is handled by the following basic fact.

Proposition A.4 *Let $\mu \in \mathcal{P}(\mathbb{R}^k)$ be a probability distribution with finite second moment. Then if $E_1, \dots, E_N \overset{i.i.d.}{\sim} \mu$ and $\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^N \delta_{E_i}$, one has*

$$\operatorname*{p-lim}_{N\to\infty} \mathbb{W}_2(\hat{\mu}_N, \mu) = 0.$$

Proof It suffices to show that $\hat{\mu}_N \to \mu$ weakly in probability and show convergence in probability of the L^2 norm. The first is clear and the second holds by the law of large numbers. □

Continuing to the inductive step, recall that the process $(U_s^{k,t})_{t\ge 1}$ is Gaussian by construction, and independent of $U_s^{k,0}$. Define

$$\begin{aligned} C_{t_1,t_2,s} &= \mathbb{E}\left[U_s^{k,t_1} U_s^{k,t_2}\right]; \\ \boldsymbol{C}_{\le t,s} &= (C_{t_1,t_2,s})_{t_1,t_2\le t}. \end{aligned}$$

We then have

$$\begin{aligned} \mathbb{E}[U_s^{k,t+1} | U_s^{k,0}, \dots, U_s^{k,t}] &= \sum_{t_1=1}^t \widetilde{\alpha}_{t_1,s} U^{k,t_1}; \\ \widetilde{\alpha}_{t_1,s} &\equiv \sum_{t_2=1}^t (\boldsymbol{C}^{-1}_{\le t,s})_{t_1,t_2} C_{t_2,t+1,s}. \end{aligned} \tag{A.55}$$

Here in writing $(\boldsymbol{C}^{-1}_{\le t,s})_{t_1,t_2}$, we view $\boldsymbol{C}_{\le t,s}$ as a $(t+1)\times(t+1)$ matrix for each $s \in \mathcal{S}$.

On the other hand, from point (a), we know that

$$\begin{aligned}\mathbb{E}[\boldsymbol{q}_s^{k,t+1}|\mathcal{F}_t] &= \sum_{1\le t_1\le t} \alpha_{t_1,s}\boldsymbol{q}_s^{t_1,k};\\ \alpha_{t_1,s} &\equiv \sum_{t_2=1}^{t} (\boldsymbol{G}^{-1}_{\xi^{k,s},t-1})_{t_1-1,t_2-1}(\boldsymbol{G}_{\xi^{k,s},t})_{t_2-1,t}.\end{aligned} \tag{A.56}$$

Moreover the induction hypothesis of (A.40) implies that for $t_1, t_2 \le t$,

$$(\boldsymbol{G}_{\xi^{k,s},t})_{t_1,t_2} \simeq \mathbb{E}\left[\xi^{k,s}\left(f_{t_1}(W_s^0,\dots,W_s^{t_1};E_s^0,\dots,E_s^{t_1}), f_t(W_s^0,\dots,W_s^{t_2};E_s^0,\dots,E_s^{t_2})\right)\right]. \tag{A.57}$$

(Recall that by definition $W_s^t \equiv \sum_{k\le D} U_s^{k,t}$, while $\boldsymbol{f}_t = f_t(\boldsymbol{V}_t;\boldsymbol{E}_t)$ here.)

Therefore, from the definition of the process $(U_s^{k,t})_{t\ge 0}$,

$$(\boldsymbol{G}_{\xi^{k,s},t})_{t_1,t_2} \simeq C_{t_1+1,t_2+1,s}, \qquad \forall t_1,t_2 \le t.$$

Recalling that $\boldsymbol{G}_{\xi^{k,s},t}$ is well-conditioned, we find (recall (A.55), (A.56)):

$$\alpha_{t_1,s} \simeq \widetilde{\alpha}_{t_1,s}.$$

Therefore we also have

$$\begin{aligned}\mathbb{E}[\boldsymbol{q}^{k,t+1}|\mathcal{F}_t] - \sum_{t_1=1}^{t}\widetilde{\alpha}_{t_1}\diamond\boldsymbol{q}^{k,t_1} &= \sum_{t_1=1}^{t}(\alpha_{t_1}-\widetilde{\alpha}_{t_1})\diamond\boldsymbol{q}^{k,t_1}\\ &\simeq 0.\end{aligned}$$

Moreover, Lemma A.1 (point 4) shows that $\mathcal{P}_t^{\perp}(\widetilde{\boldsymbol{A}}^{(k)})\{\boldsymbol{f}_t\} \simeq \widetilde{\boldsymbol{A}}^{(k)}\{(\boldsymbol{f}_t^{\otimes k-1})_{\perp}\}$ has entries which are approximately independent Gaussian with variance

$$\sigma_{t,s}^2 \equiv \left(\Gamma^{(k),\nabla}\diamond(\boldsymbol{f}_t^{\otimes k-1})_{\perp},\quad \Gamma^{(k),\nabla}\diamond(\boldsymbol{f}_t^{\otimes k-1})_{\perp}\right)$$

on coordinates $i\in\mathcal{I}_s$, even conditionally on $\mathcal{F}_t$. Therefore

$$\boldsymbol{q}^{k,t+1} \stackrel{\mathrm{d}}{=} \sum_{t_1=1}^{t}\widetilde{\alpha}_{t_1}\diamond\boldsymbol{q}^{k,t_1} + \sigma_t\diamond\boldsymbol{g} + \boldsymbol{err}^{k,t+1}, \tag{A.58}$$

where $\|\boldsymbol{err}\|_N \simeq 0$ and $\boldsymbol{g}\sim\mathsf{N}(\boldsymbol{0},\boldsymbol{I}_N)$ is independent of everything else. It now remains to verify that this agrees with the desired covariance. As proved in the previous point, for any $t_1\le t$,

$$\begin{aligned}R\left(\boldsymbol{q}^{k,t+1},\boldsymbol{q}^{k,t'+1}\right)_s &\simeq \xi^{k,s}\left(\boldsymbol{f}_t,\boldsymbol{f}_{t'}\right)\\ &\simeq \mathbb{E}\left[U_s^{k,t+1}U_s^{k,t'+1}\right].\end{aligned}$$

In particular this establishes convergence of the second moment, so in order to prove (A.40) it is sufficient to establish weak convergence. Hence we may assume $\psi:\mathbb{R}^{D\times(t+1)}\to\mathbb{R}$ is Lipschitz (rather than just pseudo-Lipschitz).

Using the representation (A.58), and focusing for simplicity on a single k, we get

$$\frac{1}{N_s}\sum_{i\in\mathcal{I}_s}\psi\left(\boldsymbol{q}_i^{k,\le t}, q_i^{k,t+1}; \boldsymbol{e}_i^{\le t}\right) \simeq \frac{1}{N_s}\sum_{i\in\mathcal{I}_s}\psi\left(\boldsymbol{q}_i^{k,\le t}, \sum_{s=1}^{t}\widetilde{\alpha}_s\boldsymbol{q}^{k,s} + \sigma_t g_i; \boldsymbol{e}_i^{\le t}\right)$$
$$\simeq \frac{1}{N_s}\sum_{i\in\mathcal{I}_s}\mathbb{E}^{g\sim\mathcal{N}(0,1)}\psi\left(\boldsymbol{q}_i^{k,\le t}, \sum_{s=1}^{t}\widetilde{\alpha}_s\boldsymbol{q}^{k,s} + \sigma_t g; \boldsymbol{e}_i^{\le t}\right).$$

The second equality above follows by Gaussian concentration since ψ is assumed Lipschitz. Applying the induction hypothesis now implies (A.40), except that $\boldsymbol{e}^{t+1}$ is not present. However since $\boldsymbol{e}_i^{t+1}$ and E_s^{t+1} have the same law and are both independent of the past, $\mathbb{W}_2$ convergence immediately transfers by Proposition A.5 below. This completes the proof of part (c).

Proposition A.5 *Let $\nu_n = \sum_{i=1}^n \delta_{\widehat{X}_i}$ for $n \ge 1$ be a sequence of probability measures on $\mathbb{R}^k$ converging to $\nu \in \mathcal{P}(\mathbb{R}^k)$ in $\mathbb{W}_2$. Let $\mu \in \mathcal{P}(\mathbb{R}^k)$ be a probability distribution with finite second moment. Let*

$$E_1, \dots, E_N \overset{i.i.d.}{\sim} \mu$$

and set

$$\widetilde{\nu}_n = \sum_{i=1}^{n}\delta_{(\widehat{X}_i, E_i)}.$$

Then

$$\operatorname*{p\text{-}lim}_{N\to\infty}\mathbb{W}_2(\widetilde{\nu}_N, \nu\otimes\mu) = 0.$$

Proof Using Proposition A.4 applied to ν, we can find for any $\varepsilon > 0$ a coupling $\Pi = \left((\widehat{X}_i, X_i)\right)_{i\in[N]}$ of ν_n with i.i.d. samples $\hat{\nu}_n$ with transport cost at most ε. Generate independent variables $E_1, \dots, E_N \overset{i.i.d.}{\sim} \mu$. Then note that

$$\mathbb{W}_2\left(\widetilde{\nu}_n, \nu\otimes\mu\right) \le \mathbb{W}_2\left(\widetilde{\nu}_n, \sum_{i=1}^{n}\delta_{(X_i,E_i)}\right) + \mathbb{W}_2\left(\sum_{i=1}^{n}\delta_{(X_i,E_i)}, \nu\otimes\mu\right)$$
$$\le \varepsilon + o_{\mathbb{P}}(1).$$

Here in the latter step we used the assumption on the coupling Π for the first term and Proposition A.4 applied to $\nu\otimes\mu$ on the second term. This completes the proof. □

A.6: Asymptotic Equivalence of AMP and Long AMP

Here we show that AMP and LAMP produce approximately the same iterates.

Lemma A.6 *Let $\{\boldsymbol{G}^{(k)}\}_{k\le D}$ be standard Gaussian tensors, and $\boldsymbol{A}^{(k)} = \Gamma^{(k)}\diamond\boldsymbol{G}^{(k)}$ for $k \ge 2$. Consider the corresponding AMP iterates $\boldsymbol{Z}_t \equiv (\boldsymbol{z}^{k,t_1})_{k\le D, t_1\le t}$ and LAMP iterates $\boldsymbol{Q}_t \equiv (\boldsymbol{q}^{k,t_1})_{k\le D, t_1\le t}$, from the same initialization $\boldsymbol{Z}_0 = \boldsymbol{Q}_0$ satisfying the assumptions of Theorems 2 and 3.*

Let $\boldsymbol{f}_t = f_t(\boldsymbol{V}_t; \boldsymbol{E}_t)$, $t \ge 0$ be the nonlinearities applied to LAMP iterates. Further assume that there exists a constant $C < \infty$ such that, for all $t \le T$,

(i) *The LAMP Gram matrices* $\boldsymbol{G}_{k,t} = \boldsymbol{G}_{k,t}$ *are well-conditioned as guaranteed by Assumption 3, i.e.,*

$$C^{-1} \leq \sigma_{\min}(\boldsymbol{G}_{k,t}) \leq \sigma_{\max}(\boldsymbol{G}_{k,t}) \leq C, \qquad \forall k \leq D,\ t \leq T.$$

(ii) *Let the linear operator* $\mathcal{T}_{k,t} : \mathbb{R}^{N\times t} \to \mathbb{R}^{N\times t}$ *be defined as per (A.22), with* $\boldsymbol{G}_{k,t} = \boldsymbol{G}_{k,t}(\boldsymbol{V})$, *and* $\boldsymbol{f}_t = f_t(\boldsymbol{V}, \boldsymbol{E}_t)$, *and define* $\mathcal{L}_{k,t} = \boldsymbol{1} + \mathcal{T}_{k,t}$. *Then*

$$C^{-1} \leq \sigma_{\min}(\mathcal{L}_{k,t}) \leq \sigma_{\max}(\mathcal{L}_{k,t}) \leq C.$$

Then, for any $t \leq T$, *we have*

$$\|\boldsymbol{Z}_t - \boldsymbol{Q}_t\|_N \simeq 0. \tag{A.59}$$

Proof Throughout the proof we will suppress $\boldsymbol{E}_t$ and simply write $f_t(\boldsymbol{W}_t)$ or $f_t(\boldsymbol{V}_t)$ to distinguish AMP and LAMP iterates, and analogously for $\boldsymbol{G}_{k,t}(\boldsymbol{W}_t)$ or $\boldsymbol{G}_{k,t}(\boldsymbol{V}_t)$. The proof is by induction over the iteration number, so we will assume it to hold at iteration t, and prove it for iteration $t+1$. We prove the induction step by establishing the following two facts for each $2 \leq k \leq D$:

$$\left\|\mathsf{AMP}_{t+1}(\boldsymbol{Z}_t)_k - \mathsf{AMP}_{t+1}(\boldsymbol{Q}_t)_k\right\|_N \simeq 0, \tag{A.60}$$

$$\left\|\mathsf{AMP}_{t+1}(\boldsymbol{Q}_t)_k - \mathsf{LAMP}_{t+1}(\boldsymbol{Q}_t)_k\right\|_N \simeq 0. \tag{A.61}$$

Let us first consider the claim (A.60), and note that

$$\begin{aligned}\mathsf{AMP}_{t+1}(\boldsymbol{Z}_t)_k - \mathsf{AMP}_{t+1}(\boldsymbol{Q}_t)_k = {} & \boldsymbol{A}^{(k)}\{f_t(\boldsymbol{W}_t)\} - \boldsymbol{A}^{(k)}\{f_t(\boldsymbol{V}_t)\} \\ & - \sum_{t_1 \leq t} d_{t,t_1,k} \diamond \big(f_{t_1-1}(\boldsymbol{W}_{t_1-1}) - f_{t_1-1}(\boldsymbol{V}_{t_1-1})\big),\end{aligned}$$

where we wrote $d_{t,t_1,k,s}$ for the coefficients of (A.8), with AMP iterates replaced by LAMP iterates. We then have

$$\begin{aligned}\left\|\mathsf{AMP}_{t+1}(\boldsymbol{Z}_t)_k - \mathsf{AMP}_{t+1}(\boldsymbol{Q}_t)_k\right\|_N &\leq D_{1,t} + D_{2,t}; \\ D_{1,t} &\equiv \left\|\boldsymbol{A}^{(k)}\{f_t(\boldsymbol{W}_t)\} - \boldsymbol{A}^{(k)}\{f_t(\boldsymbol{V}_t)\}\right\|_N, \\ D_{2,t} &\equiv \sum_{t_1 \leq t,\ s\in\mathcal{S}} |d_{t,t_1,k,s}| \cdot \left\|f_{t_1-1,s}(\boldsymbol{W}_{t_1-1}) - f_{t_1-1,s}(\boldsymbol{V}_{t_1-1})\right\|_N.\end{aligned}$$

Notice that, by the induction assumption (and recalling that each $f_{t,s}$ is Lipschitz continuous and acts component-wise):

$$\left\|f_t(\boldsymbol{W}_t) - f_t(\boldsymbol{V}_t)\right\|_N \leq C_T \sum_{t_1 \leq t,\ k\leq D} \|\boldsymbol{w}^{k,t_1} - \boldsymbol{v}^{k,t_1}\|_N \simeq 0. \tag{A.62}$$

Further, for any tensor $\boldsymbol{T} \in (\mathbb{R}^N)^{\otimes k}$, and any vectors $\boldsymbol{v}_1, bv_2 \in \mathbb{R}^N$,

$$\|\boldsymbol{T}\{\boldsymbol{v}_1\} - \boldsymbol{T}\{\boldsymbol{v}_2\}\|_N \leq (N^{\frac{k-2}{2}}\|\boldsymbol{T}\|_{\text{op}})(\|\boldsymbol{v}_1\|_N + \|\boldsymbol{v}_2\|_N)^{k-2}\|\boldsymbol{v}_1 - \boldsymbol{v}_2\|_N \tag{A.63}$$

Using Lemma A.1, this implies that the following bound holds with high probability for a constant C:

$$\begin{aligned}D_{1,t} &\leq C(\|f_t(\boldsymbol{W}_t)\|_N + \|f_t(\boldsymbol{V}_t)\|_N)^{k-2}\|f_t(\boldsymbol{W}_t) - f_t(\boldsymbol{V}_t)\|_N \\ &\leq C(2\|f_t(\boldsymbol{V}_t)\|_N + \|f_t(\boldsymbol{W}_t) - f_t(\boldsymbol{V}_t)\|_N)^{k-2}\|f_t(\boldsymbol{W}_t) - f_t(\boldsymbol{V}_t)\|_N \\ &\simeq 0.\end{aligned}$$

The last step follows from (A.62) and Theorem 3, which implies (recall each $f_{t,s}$ is Lipschitz) that $\|f_t(\boldsymbol{V}_t)\|_N \le C$ with probability $1 - o(1)$. Notice that the same argument implies $\|f_t(\boldsymbol{W}_t)\|_N \le C$ with high probability.

Similarly, $D_{2,t} \simeq 0$ follows since $\|f_{t_1-1}(\boldsymbol{W}_{t_1-1}) - f_{t_1-1}(\boldsymbol{V}_{t_1-1})\|_N \simeq 0$ and $|d_{t,t_1,k,s}| \le C_T$ by construction, thus yielding (A.60).

We now prove (A.61). Comparing (A.8) and (A.29), with $\mathcal{P}_t^{\|} = \mathbf{1} - \mathcal{P}_t^{\perp}$ we find

$$\mathsf{AMP}_{t+1}(\boldsymbol{Q}_t)_k - \mathsf{LAMP}_{t+1}(\boldsymbol{Q}_t)_k = \mathcal{P}_t^{\|}(\boldsymbol{A}^{(k)})\{f_t(\boldsymbol{V}_t)\} - \mathbf{ons}_{k,t+1} - \sum_{0 \le t_1 \le t-1} h_{t,t_1,k} \diamond \boldsymbol{q}^{k,t_1+1},$$

$$\mathbf{ons}_{k,t+1} = \sum_{t_1 \le t} d_{t,t_1,k} \diamond f_{t_1-1}(\boldsymbol{V}_{t_1-1}) \tag{A.64}$$

Note that $\mathcal{P}_t^{\|}(\boldsymbol{A}^{(k)}) = \mathbb{E}\left[\boldsymbol{A}^{(k)} | \mathcal{F}_t\right]$, where $\mathcal{F}_t$ here is the analogous σ-algebra generated by $\{\mathbf{q}^{k,t_1}, \boldsymbol{e}^{t_1}\}_{t_1 \le t, k \le D}$. Equivalently, this is the conditional expectation of $\boldsymbol{A}^{(k)}$ given the linear constraints

$$\boldsymbol{A}^{(k)}\{f_{t_1}(\boldsymbol{V}_{t_1})\} = \boldsymbol{y}_{k,t_1+1}, \quad \text{for } t_1 \in \{0, \dots, t-1\}, \tag{A.65}$$

Also notice that, by the induction hypothesis, and the definition of $\boldsymbol{y}_{k,t_1}$, (A.14), we have for all $t_1 \le t$,

$$\boldsymbol{y}_{k,t_1} \simeq \boldsymbol{q}^{k,t_1} + \mathbf{ons}_{k,t_1}. \tag{A.66}$$

Lemma A.2 implies that $\mathcal{P}_t^{\|}(\boldsymbol{A}^{(k)})$ takes the form of (A.21) for a suitable matrix $\widehat{\boldsymbol{Z}}_{k,t} \in \mathbb{R}^{N \times t}$. The key claim is that

$$\widehat{\boldsymbol{Z}}_{k,t} \simeq \boldsymbol{Q}_t. \tag{A.67}$$

In order to establish this claim, we show that, under the inductive hypothesis,

$$(\mathbf{1} + \mathcal{T}_{k,t})\, \boldsymbol{Q}_t \simeq \boldsymbol{Y}_{k,t}. \tag{A.68}$$

Since $\mathcal{L}_{k,t} = \mathbf{1} + \mathcal{T}_{k,t}$ is well-conditioned by assumption, the combination of (A.23) and (A.68) implies $\widehat{\boldsymbol{Z}}_{k,t} \simeq \boldsymbol{Q}_t$. By (A.66), in order to prove (A.68), it is sufficient to show that

$$\mathcal{T}_t\, \boldsymbol{Q}_t \simeq \mathbf{ONS}_{k,t} \equiv [\mathbf{ons}_{k,1} | \cdots | \mathbf{ons}_{k,t}]. \tag{A.69}$$

In order to prove (A.69), we use Theorem 3. Recall that

$$C_{t_1,t_2,s} = \mathbb{E}\{U_s^{k,t_1} U_s^{k,t_2}\},$$
$$W_s^{t_1} = \sum_{2 \le k \le D} U_s^{k,t_1},$$
$$\boldsymbol{C}_{\le t} = (C_{t_1,t_2,s})_{t_1,t_2 \le t}.$$

(The value $2 \le k \le D$ is implicitly fixed in the definition of $\boldsymbol{C}_{\le t}$.) By Theorem 3,

$$C_{t_1+1,t_2+1} \simeq \langle \boldsymbol{q}^{k,t_1+1}, \boldsymbol{q}^{k,t_2+1} \rangle \simeq (\boldsymbol{G}_{\xi^{k,s},t}(\boldsymbol{V}))_{t_1,t_2}, \quad \forall\, t_1, t_2 \le t.$$

This implies for any $0 \le t_1 \le t-1$ and $s \in \mathcal{S}$,

$$\begin{aligned}
&\sum_{t_2=0}^{t-1} (\boldsymbol{G}^{-1}_{\xi^{k,s},t-1})_{t_1,t_2} R\left(\boldsymbol{q}^{k,t_2+1}, f_{t-1}(\boldsymbol{V}_{t-1})\right)_s \\
&\simeq \sum_{t_2=0}^{t-1} (\boldsymbol{C}^{-1}_{\le t,s})_{t_1+1,t_2+1}\, \mathbb{E}\left[U_s^{k,t_2+1} f_{t-1,s}(W_s^0,\dots,W_s^{t-1};E_s^0,\dots,E_s^{t-1})\right] \\
&= \mathbb{E}\left[\frac{\partial f_{t-1,s}}{\partial W_s^{t_1+1}}(W_s^0,\dots,W_s^{t-1};E_s^0,\dots,E_s^{t-1})\right]\mathbf{1}_{t_1\le t-2}. \qquad \text{(A.70)}
\end{aligned}$$

Indeed, Gaussian integration by parts yields the latter expression (it can be done conditionally on the variables E since they are independent). Combining (A.70) with the definition (A.8) will now allow us to conclude $\mathcal{T}_{k,t}\,\boldsymbol{Q}_t \simeq \mathbf{ONS}_{k,t}$ as desired. Indeed for each $s \in \mathcal{S}$ we have

$$\begin{aligned}
\left[\mathcal{T}_{k,t}\,\boldsymbol{Q}_t\right]_{t,s} &= \sum_{t_1=0}^{t-1} \partial_{s'}\xi^{k,s}(\vec{R}(\boldsymbol{f}_{t_1},\boldsymbol{f}_{t-1}))\Big(\sum_{t_2=0}^{t-1}(\boldsymbol{G}^{-1}_{\xi^{k,s'},t-1})_{t_1,t_2}\quad R_{s'}(\boldsymbol{q}^{k,t_2+1},\boldsymbol{f}_{t-1})\Big)\boldsymbol{f}_{t_1,s} \\
&\simeq \sum_{t_1=0}^{t-2}\sum_{s'\in\mathcal{S}} \partial_{s'}\xi^{k,s}\big(\vec{R}(\boldsymbol{f}_{t_1},\boldsymbol{f}_{t-1})\big)\cdot\mathbb{E}\left[\frac{\partial f_{t-1,s'}}{\partial W_{s'}^{t_1+1}}(W_{s'}^0,\dots,W_{s'}^{t-1})\right]\boldsymbol{f}_{t_1,s} \\
&= \mathbf{ons}_{k,t}.
\end{aligned}$$

Having established (A.67), we now use the formula (A.21) for $\mathcal{P}_t^{\|}(\boldsymbol{A}^{(k)}) = \mathbb{E}\left[\boldsymbol{A}^{(k)}|\mathcal{F}_t\right]$. The result is:

$$\begin{aligned}
\mathcal{P}_t^{\|}(\boldsymbol{A}^{(k)})\{\boldsymbol{f}_t\} &\simeq \sum_{t_1\le t}\big(\alpha_{t_1}\diamond\boldsymbol{q}^{k,t_1}+\beta_{t_1}\diamond\boldsymbol{f}_{t_1}\big); \\
\alpha_{t_1,s} &\equiv \sum_{0\le t_2\le t-1}(\boldsymbol{G}^{-1}_{\xi^{k,s},t-1})_{t_1,t_2}\;\xi^{k,s}\big(\vec{R}(f_{t_2}(\boldsymbol{V}_{t_2}),f_t(\boldsymbol{V}_t))\big), \\
\beta_{t_1,s} &\equiv \sum_{s'\in\mathcal{S}}\partial_{s'}\xi^{k,s}\big(\vec{R}(\boldsymbol{f}_{t_1},\boldsymbol{f}_t)\big)\left(\sum_{0\le t_2\le t-1}(\boldsymbol{G}^{-1}_{\xi^{k,s'},t-1})_{t_1,t_2}\;R_{s'}(\boldsymbol{q}^{k,t_2},\boldsymbol{f}_t)\right).
\end{aligned} \qquad \text{(A.71)}$$

On the other hand, using again (A.70) gives

$$\begin{aligned}
\sum_{t_1\le t}\beta_{t_1}\diamond\boldsymbol{f}_{t_1} &\simeq \sum_{t_1\le t-1} d_{t,t_1,k}\diamond\boldsymbol{f}_{t_1-1} = \mathbf{ons}_{k,t+1}, \\
\sum_{t_1\le t}\alpha_{t_1}\diamond\boldsymbol{q}^{k,t_1} &\simeq \sum_{0\le t_1\le t-1} h_{t,t_1,k}\diamond\boldsymbol{q}^{k,t_1+1}.
\end{aligned}$$

We conclude from (A.64) that $\|\mathsf{AMP}_{t+1}(\boldsymbol{Q}_t)_k - \mathsf{LAMP}_{t+1}(\boldsymbol{Q}_t)_k\|_N \simeq 0$. This concludes the proof. □

References

1. Auffinger, A., Ben Arous, G., Černý, J.: Random matrices and complexity of spin glasses. Commun. Pure Appl. Math. **66**(2), 165–201 (2013)

2. Auffinger, A., Chen, W.-K.: Free energy and complexity of spherical bipartite models. J. Stat. Phys. **157**(1), 40–59 (2014)
3. El Alaoui, A., Montanari, A.: Algorithmic thresholds in mean field spin glasses. arXiv preprint (2020). arXiv:2009.11481
4. El Alaoui, A., Montanari, A., Sellke, M.: Optimization of mean-field spin glasses. Ann. Probab. **49**(6), 2922–2960 (2021)
5. El Alaoui, A., Sellke, M.: Algorithmic pure states for the negative spherical perceptron. J. Stat. Phys. **189**(2), 27 (2022)
6. Bayati, M., Lelarge, M., Montanari, A.: Universality in polytope phase transitions and message passing algorithms. Ann. Appl. Probab. **25**(2), 753–822 (2015)
7. Bayati, M., Montanari, A.: The dynamics of message passing on dense graphs, with applications to compressed sensing. IEEE Trans. Inf. Theory **57**, 764–785 (2011)
8. Berthier, R., Montanari, A., Nguyen, P.-M.: State evolution for approximate message passing with non-separable functions. Inf. Inference **9**, 33–79 (2019)
9. Bolthausen, E.: An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. Commun. Math. Phys. **325**(1), 333–366 (2014)
10. Chen, W.-K., Lam, W.-K.: Universality of approximate message passing algorithms. Electron. J. Probab. **26**, 1–44 (2021)
11. Dudeja, R., Lu, Y.M., Sen, S.: Universality of approximate message passing with semi-random matrices. Ann. Probab. **51**(5), 1616–1683 (2023). https://doi.org/10.1214/23-AOP1628
12. Dembo, A., Montanari, A., Sen, S.: Extremal cuts of sparse random graphs. Ann. Probab. **45**(2), 1190–1217 (2017)
13. Fan, Z.: Approximate message passing algorithms for rotationally invariant matrices. Ann. Stat. **50**(1), 197–224 (2022)
14. Feng, O.Y., Venkataramanan, R., Rush, C., Samworth, R.J., et al.: A unifying tutorial on approximate message passing. Found. Trends Mach. Learn. **15**(4), 335–536 (2022)
15. Huang, B., Sellke, M.: Tight Lipschitz hardness for optimizing mean field spin glasses. arXiv preprint (2021). arXiv:2110.07847
16. Huang, B., Sellke, M.: Algorithmic threshold for multi-species spherical spin glasses. arXiv preprint (2023). arXiv:2303.12172
17. Huang, B., Sellke, M.: Strong topological trivialization of multi-species spherical spin glasses. arXiv preprint (2023). arXiv:2308.09677
18. Javanmard, A., Montanari, A.: State evolution for general approximate message passing algorithms, with applications to spatial coupling. Inf. Inference **2**(2), 115–144 (2013)
19. Krzakala, F., Montanari, A., Ricci-Tersenghi, F., Semerjian, G., Zdeborová, L.: Gibbs states and the set of solutions of random constraint satisfaction problems. Proc. Natl. Acad. Sci. **104**(25), 10318–10323 (2007)
20. McKenna, B.: Complexity of bipartite spherical spin glasses. arXiv preprint (2021). arXiv:2105.05043
21. Montanari, A.: Optimization of the Sherrington–Kirkpatrick Hamiltonian. SIAM J. Comput. (2021). https://doi.org/10.1137/20M132016X
22. Panchenko, D.: On the K-sat model with large number of clauses. Random Structures & Algorithms **52**(3), 536–542 (2018)
23. Richard, E., Montanari, A.: A statistical model for tensor PCA. In: Advances in Neural Information Processing Systems, pp. 2897–2905 (2014)
24. Sellke, M.: Optimizing mean field spin glasses with external field. Electron. J. Probab. **29**, 1–47 (2024)
25. Sherrington, D., Kirkpatrick, S.: Solvable model of a spin-glass. Phys. Rev. Lett. **35**(26), 1792 (1975)
26. Subag, E.: Following the ground states of full-RSB spherical spin glasses. Commun. Pure Appl. Math. **74**(5), 1021–1044 (2021)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.